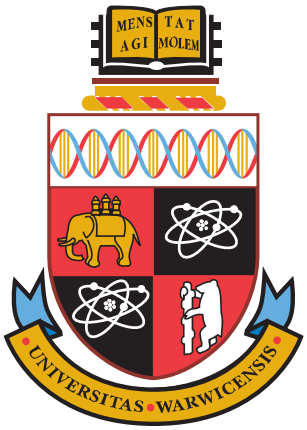


The Emotional Recall Task

By

M a s i t a h

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Psychology

December 2018

THE UNIVERSITY OF
WARWICK

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Bismillahi rahmanirrahim…

Alhamdulillahirabbilalamin. Upon the grace of Allah Subhanahu Wata 'ala, the thesis entitled "The Emotional Recall Task" can be completed properly.

Undertaking this PhD has been a truly life-changing experience. It would not have been possible for me to overcome all the obstacles and barriers without help and support from others. For it, I would like to express my sincere gratitude to my supervisor Professor Thomas Hills for his guidance, great support, prompt help, and kind advice throughout my PhD journey. Such a real privilege and an honour for me to share of his exceptional scientific knowledge on study ideas, strategic career choices, and also his extraordinary human qualities.

I would also like to thank Professor Gordon Brown, as my second supervisor, for his support and feedback during my PhD upgrade.

For her kindness and support as a mentor, as well as a friend, I would like particularly to thank Dr. Elisabeth Blagrove for supporting me on attending the Association for Psychological Science (APS) Conference. Participating in this conference brought me an opportunity to become the APS Campus Representative 2018.

For all the progressive feedbacks and supports during my annual review, I would like to thank Professor Elizabeth Maylor, Dr. Sakari Lemola, Dr. Jesse Preston, and Dr. Claudie Fox.

For all the fruitful discussion during the language group meetings, I would also like to thank Professor Sotaro Kita and Dr. Suzanne Aussems.

I would also like to acknowledge Professor Anu Realo and Professor Laura Mickes - my viva examiners, for their very helpful comments and suggestions which helped me to improve the presentation of this thesis.

My fellowship was funded by Indonesia Endowment Fund for Education (IEFE), Ministry of Finance of Republic of Indonesia. I appreciate the council for seeing potential in the proposed work and allowing me to carry it out successfully.

The Department of Psychology at the University of Warwick has been my life during the last four years. I have enjoyed a wonderful time and am very grateful to have had this opportunity.

I would also like to thank the Indonesian Student Association Community in the United Kingdom for their friendship and the warmth during my time in the UK, and for always making me feel so welcome.

For all the knowledge-sharing and support during the PhD application process, I would like to show my gratitude to Indonesian psychologists of University of Indonesia and University of Sumatera Utara. Especially to Dr. Zulkarnain Amin who support me to conduct the ERT Cultural Studies in Indonesia.

Furthermore, I would like to thank the research participants, collaborators, and data collection assistants who took part in the ERT research reported in this dissertation. Without them, this research would not be possible.

Finally, and most importantly, I would like to thank my parents Dr. M. Ali Musri S. and Professor Julia Reveny.

DECLARATIONS

This thesis is submitted to the University of Warwick in support of the application for the degree of Doctor of Philosophy. It has been composed by the author and has not been submitted in any previous application for any degree.

The work presented (including data collection and data analyses) was carried out by the author. Parts of this thesis have been submitted or prepared for publication by the author. My supervisor, Thomas Hills, supported my research throughout and provided feedback on the writing.

ABSTRACT

Existing self-report affect scales typically involve recognition of emotions from a predetermined emotion checklist. Therefore, we propose an emotion scale that relies on recalled memory. Instead of asking people to evaluate their emotional experience in relation to a list of terms that may nor may not properly cover their entire emotion space, we ask people to produce 10 words that best describe their past emotions and then to rate how often they have experienced these emotions. The Emotion Recall Task (ERT) approach leverages on the more effortful and accurate retrieval and recognition processes and avoids many concerns that surround recognition-based affect scales such as emotional breadth and specificity. In comparisons with the PANAS, arguably the most commonly used affect scale, the ERT performs at least equally well in predicting related constructs such as well-being, life satisfaction and depression. In a test-retest validation study, the ERT is found to be a reliable instrument for deriving recalled emotion with correlations on par with other existing affect scales. This thesis further presents the ERT 2.0, which was constructed by allowing participants to rate their own valence and arousal on each emotion. The development of the ERT 2.0 increases the accessibility to conduct the ERT studies in different cultures and languages. To provide insight into the cognitive processes underlying emotional recall, a series of formal interviews with participants following the ERT revealed that most participants start by retrieving situational events they recently experienced and then select experiences that evoked specific emotions. Less frequently, people retrieved emotions first. This shows some distinct similar with other forms of semantic memory search. To provide a demonstration of the cultural validity of the ERT, a cultural project a cultural project was undertaken to compare Indonesian participants with American participants. The ERT result for Indonesian emotional expression provided a new insight contrasted with the established culture ideas, revealing that Indonesians' recall-based affect may be less extreme and more negative than previously thought, despite Indonesians and Americans appearing to draw from a

similar set of emotional words. In a further comparison, we developed *the emotional literacy test* ("name all the emotions you think of") to understand how variability in the emotional lexicon influenced recalled emotions. The ERT was then compared with an emotional heuristic evaluation – the peak-end rule. Both peak and end emotions reflected a positive relationship with the ERT: peak and end emotions are both significantly predict independent components of the ERT variance and together explain approximately half of the variance of the ERT. The final chapter describes on how one's own emotional lexicon influences the emotions that one recalls. By asking people listing all the emotions they can think of, we assess the extent to which their emotional lexicon can predict their ERT. To measure each individual's emotional lexicon, we introduce emotional literacy.

Chapter 1

Introduction

The aim of this research is to investigate a new method for evaluating emotions based on emotional recall. The hope is that this will provide insight into the investigation of emotional memory, allow us to evaluate the differences between emotional recall and recognition—as all existing scales are based on recognition—and to understand the relationship between search of emotional memory and existing measures of affect. This chapter will provide an exploration of the theory needed in understanding discrete emotion, the description of some of the most common emotional scales (also called affect scales), and then provide a brief introduction about the Emotional Recall Task, which this thesis develops over a series of studies.

1.1. Emotional Dimensionality

The consensus of discrete emotion theory argues that there is a small number of core emotions. For example, there is Tomkins (1984) nine innate affect theory (the positive affect: *interest/excitement, enjoyment, surprise/startle; the negative affect: distress/anguish, fear/terror, shame/humiliation, contempt, disgust, anger/rage*) and Ekman's (1992) six basic emotion theory (*anger, disgust, fear, happiness, sadness, surprise*). Further exploration of emotions examined how affective reactions arise from cognitive response about actions, events, or objects (Clore, Schwarz, & Conway, 1994). The experimental test conducted by Roseman (1991) explained how combinations of five determinant factors decide people's qualitatively different emotional responses. These determinant factors are motivational state (rewarding or punishing), situational state (present or

absent), probability (certain or uncertain), legitimacy (deserved positive or negative outcome), and causal agency (such as other person or self).

The history of speculations about human emotions dates back to at least Aristotle's *Nicomachean Ethics* (Broadie & Rowe, 2002), which lists 11 different emotions, including 'pity' and 'emulation' (the act of copying another individual's behaviour). Darwin (1872), taking an evolutionary approach, attempted to classify emotions in relation to their adaptive value, and in addition to high and low valence emotions, included such dimensions as 'surprise', 'meditation' and 'shyness'. Looking across cultures, Ekman (1992) proposed a set of 'natural kinds' for emotions, which included *happiness, surprise*, *fear, sadness, anger, and disgust.*

A dimensional approach to developing a basic description of personal emotions as it is accessible through introspection was first described by Wundt (1905a). The three separate dimensions Wundt (1905b) proposed were valence (positive-negative), arousal (calm-excited), and tension (tense-relaxed). The tension dimension, however, is in practice often difficult to distinguish from arousal. Therefore, modern approaches have focused primarily on the valence and arousal dimensions. On the other hand, according to Scherer (2005), the major drawbacks of this approach are the difficulty of knowing whether the valence dimension describes the intrinsic quality of the feeling.

Representation of an emotional experience may nonetheless reflect individual differences in how people understand their affective experiences (for reviews see Barrett, Gross, Christensen, & Benvenuto, 2001; Clore, Gasper, Garvin, & Forgas, 2001; Clore & Tamir, 2002). One way this has been characterized is in relation to emotional granularity (for reviews see Barrett, 1998, 2004; Barrett et al., 2001; Feldman, 1995).

People with high emotional granularity report their emotional experience in differentiated terms with discrete emotion labels to capture their distinctiveness, like *joy* instead of *happy* and are also more likely to recognize the subtleties of various emotional dimensions. Individuals with less granularity describe their emotional experience using more general emotional labels to reveal their affect and focus more exclusively on valence (for reviews of core affect, see Russell, 2003; Russell & Barrett, 1999; Tugade, Fredrickson, & Feldman Barrett, 2004).

Individual differences in affective experiences. Previous work suggests that affective experiences stimulate a certain affect-relevant concept that can be used to understand affective reactions. However, affective reactions do not spontaneously activate memories of other similarly-valenced semantic concepts (Wyer, Clore, & Isbell, 1999).

Previous work suggests that affect is elicited by reflecting on specific events and thoughts about a past experience. For example, research from Lazarus (1982) correlates affect with our cognitive appraisal of a situation. If the affect resulted from cognitive appraisal is emotionally intense, it may stimulate an attempt to explain its occurrence. As a result, the concepts and knowledge that are involved in this cognitive activity become more accessible in memory.

Other cognitive psychologists have suggested that exposure to a stimulus may automatically produce an affective categorization process. This categorization may then be available in future recall. According to Bower (1981), people with positive affective states should tend to recall and use positively-valenced concepts and knowledge whereas individuals with negative affective state should be more disposed to recall and negatively-valenced memories. It is not, however, well understood whether positive and negative affect influence the recall process or whether are the products of the recall process. Are people sad because they recall sad memories or do they recall sad memories because they are sad?

This naturally leads to the question of how affective memories and potentially affective states may differ depending on how affective memory is searched. Psychologists differentiate two types of memory retrieval: recall and recognition (Tulving, 1985). Unfortunately, compared with recognition, recall has received little attention in emotional scales. The next section provides detailed information about recall and recognition followed by a review of the most common emotional self-report scales.

1.2. Recall and Recognition: similarities and differences

Language is a way to express people's emotion. It is used to show the reflective meaning of emotional responses to an experience, by combining recall and recognised emotions. As one of three core processes of memory, along with

encoding and storage, recall can be explained as the mental process to retrieve a stimulus from memory based on an associated memory probe (Atkinson & Shiffrin, 1968). Recognition refers to the ability to recognise the presence of a stimuli in memory (e.g., the presence of a probe). Recognition memory can be divided up into recollection processes and familiarity processes.

Several theorists (Anderson & Bower, 1972; Bahrick, 1970; Estes & DaPolito, 1967; Kintsch, 1970) proposed that a principal difference between recall and recognition is the retrieval stage of memory. In some cases, recognition has been claimed to be superior to recall tests because less information is needed to pass a recognition test (Postman, 1963). Recall involves a more comprehensive process (Watkins & Gardiner, 1979). To recall emotions correctly, an item must be stored in memory in sufficient detail to be recovered based on a generic probe (e.g., 'recall an emotion'). Recognition, on the other hand, enables a response based on all or part of a stimulus that is already present in memory (Hollingworth, 1913) and therefore does not require a search of memory.

The decisive role of memory search is evidenced by the observation that much more information is available in memory than is accessible at any moment. In one experiment (Tulving & Pearlstone, 1966), participants learned various lists of word on a single trial. Immediately after presenting the lists, the experimenter instructed the participants to remember as many words as possible with two recall tests in succession. Three independent variables were manipulated. (a) list length (12, 24, and 48 words), (b) number of words or item per category (1, 2, and 4 words), and (c) conditions of recall in the first recall test (cued recall and free recall). The second recall test was given under the condition of cued recall. Results showed cued recall was higher than free recall. This differentiation varied directly with list length and inversely with number of items per category. This finding indicated that sufficiently intact memory traces of many words not recalled under the free recall conditions were available in memory, but not accessible given a generic instruction to probe memory. The discrepancy between the availability of information and its accessibility to consciousness affirms the critical role of retrieval processes.

Several additional studies illustrate the influence of memory search on recalling stored information. During three successive word recall tests (Tulving,

1967), only 50% of the words were recalled on all three tests. Participants recalled words on the second test that they missed on the first test. Another study (Williams & Hollan, 1981) had subjects spend one hour daily to recall the names of people from their high school, and found that subjects recalled a new name after as much as 10 hours in the experiment. The new names must have been available even in the first hour of the experiment, but they were not yet accessible. Regarding both studies, Koriat (2000) raises several questions: First, what prevented all the items being recalled at the beginning of testing? Second, what allowed them to become accessible later on? And third, what is the process by which people search for and recollect stored information in long term memory?

The above results are all suggestive of differences between recall and recognition. These seem especially vital in relation to emotions. In particular, is general affect more likely the result of recalled emotions or recognized emotions? We may ask ourselves how we are feeling and remember that we were happy earlier that day. Rarely do we present ourselves with a list of emotions and ask which of these we felt and how much. Rather, we recall specific emotions, potentially based on their association with specific events. However, all known emotional scales involve recognition of emotions that may or may not match the kinds of emotions we often recall. This may mean that existing emotional scales based on recognition are less valid than a scale based on recall. Before we present the Emotional Recall Test, which explores a method for assessing recall-based emotions, we first provide a brief outline of the most commonly used emotional scales.

1.3. Common Emotional Scales

Self-report is the most popular method to measure emotions (Diener, Napa-Scollon, Oishi, Dzokoto, & Suh, 2000; Robinson & Clore, 2002; Watson, 2000). Common self-report emotional scales are described below. Importantly, all of these are based more explicitly on recognition-based evaluations of emotional states.

a. Affect Balance Scale

The Affect Balance Scale (ABS: Bradburn, 1969), also known as Bradburn Scale of Psychological Well-being, was established in the early 1960s by the National Opinion Research Center. The ABS comprised of two independent dimensional models of psychological well-being; negative affect and positive affect. Each component consists of five "Yes" or "No" items. This scale asks participants: *During the past few weeks (did you feel)* .... Participants receive 1 point for each "Yes" response in positive and negative affect subscales. The overall balance score, also known as an affect difference score, is calculated by subtracting the negative affect score from the positive affect score.

Considerable literature has been devoted to explaining the meaning psychological wellbeing. However, Ryff (1989) claimed that the reigning measures of the ABS have little theoretical grounding and failed to represent key aspects of positive functioning (i.e., self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life, and personal growth). However, design of the scale as a whole shows advantages over alternative instruments as an indicator of emotional wellbeing.

ABS and its subscales were validated through groups analysis and independent criteria. Comparisons of mean scores of 2 elderly samples, 8 psychiatric outpatients, and 19 normal subjects, indicate significant discriminability of these scales. The subscales were not significantly correlated (r=-.32; p>.05), indicating that PAS and NAS are considered to be relatively independent. PAS correlated positively with morale and avowed happiness, while NAS correlated with poorer mental health and the greater role loss.

Directions: Now, let's talk about something else. We are interested in the way people are feeling these days. Looking at your present life situation, have you ever felt:

		Yes	No
* 1.	Particularly excited or interested in something?	—	—
† 2.	So restless you couldn't sit long in a chair?	—	—
* 3.	Proud because someone complimented you on something you had done?	—	—
† 4.	Very lonely or remote from other people?	—	—
* 5.	Pleased about having accomplished something?	—	—
† 6.	Bored?	—	—
* 7.	On top of the world?	—	—
† 8.	Depressed or very unhappy?	—	—
* 9.	That things were going your way?	—	—
†10.	Upset because someone criticized you?	—	—

*PAS items †NAS items

Figure 1.1. The Bradburn's affect balance scale

b. Profile of Mood States

The profile of mood states (POMS: McNair, Lorr, & Droppleman, 1981) is a list of adjectives rated on a 5-point scale commonly used to measure psychological distress. Each adjective in the POMS is scored ranging from *not at all* (0) to *extremely* (4), except *relaxed* and *efficient* are scored with reverse scaling ranging from *not at all* (4) to *extremely* (0). POMS score represents six factors that can be calculated into a Total Mood Disturbance (TMD) score as noted below.

TMD = (Tension + Depression + Anger + Fatigue + Confusion) – Vigour

The total score for *tension* (9 items) is determined by adding the scores for: *tense, shaky, on edge, panicky, relaxed, uneasy, restless, nervous,* and *anxious*.

The total score for *depression* (15 items) is determined by adding the scores for: *unhappy, sorry for things done, sad, blue, hopeless, unworthy, discouraged, lonely, miserable, gloomy, desperate, helpless, worthless, terrified,* and *guilty.*

The total score for *anger* (12 items) is determined by adding the scores for: *anger, peeved, grouchy, spiteful, annoyed, resentful, bitter, ready to fight, rebellious, deceived, furious,* and *bad tempered.*

The total score for *fatigue* (7 items) is determined by adding the scores for: *worn out, listless, fatigued, exhausted, sluggish, weary,* and *bushed*.

The total score for *confusion* (7 items) is determined by adding the scores for: *confused, unable to concentrate, muddled, bewildered, efficient, forgetful,* and *uncertain about things*.

The total score for *vigour* (8 items) is determined by adding the scores for: *lively, active, energetic, cheerful, alert, full of pep, carefree,* and *vigorous*.

The following adjectives are not used in the scoring (dummy items): *friendly, clear headed, considerate, sympathetic, helpful, good natured,* and *trusting*.

The 65-items POMS usually takes between 3 and 7 minutes. However, a few shortened forms developed to facilitate the use of POMS with under stress or pain patients; 37-item POMS (Shacham, 1983); 11-item Total Mood Disturbance Score of the POMS (TMDS-POMS; Cella et al., 1987); and 24-item Profile of Mood States – Adolescents (POMS-A; Terry, Lane, Lane, & Keohane, 1999).

c. Scale of Positive and Negative Experience

Diener et al. (2010) created the Scale of Positive and Negative Experience (SPANE), which incorporates broad descriptors for positive and negative emotions. SPANE score ranging from *very rarely or never* (1) to *very often or always* (5) based on the frequency of emotions during the past month. The measure can be used to derive an overall emotion balance score but can also be divided into positive and negative emotion scales.

Positive emotions (SPANE-P): add the scores, varying from 1 to 5, for the six items: *positive, good, pleasant, happy, joyful,* and *contented*. The score can vary from 6 (lowest possible) to 30 (highest positive emotions score).

Negative emotions (SPANE-N): add the scores, varying from 1 to 5, for the six items: *negative, bad, unpleasant, sad, afraid,* and *angry*. The score can vary from 6 (lowest possible) to 30 (highest negative emotions score).

Emotion Balance (SPANE-B): the negative emotions score is subtracted from the positive emotions score, and the resultant difference score can vary from -24 (unhappiest possible) to 24 (highest emotions balance possible). Participants with a score of 24 report that they rarely or never experience any negative emotions, and very often experience all of the positive emotions.

According to Diener et al. (2010), SPANE-P produced one strong factor with an eigenvalue above 1.0 (3.69), accounting for 61 percent of the variance in the scale items. The loadings varied from .58 to .81. The SPANE-N had one strong eigenvalue above one (3.19) that accounted for 53 percent of the variance in the scale. The factor loadings varied from .49 to .78. SPANE-P and SPANE-N correlated r=-.60 (N=682, p=.001) with each other.

Table 1.1. SPANE correlations with PANAS

	SPANE-P	SPANE-N	SPANE-B	PANAS-PA
PANAS-PA	.61	-.44	.58	
	N=505	N=499	N=499	
PANAS-NA	-.46	.70	-.65	-.31
	N=504	N=498	N=498	N=502

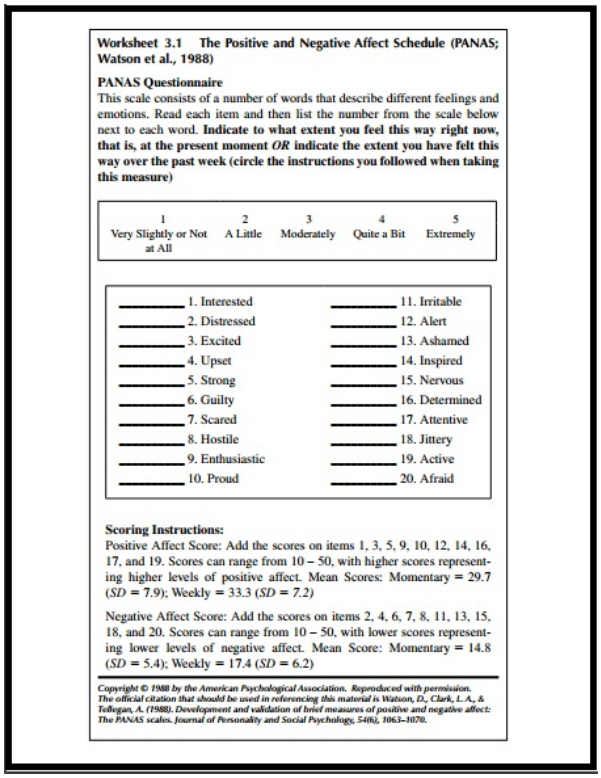
All p's<.001

Diener et al. (2010) proposed SPANE as an improvement on existing measures of emotions and a resolution for the issue with the most popular current scale of the emotions, the PANAS. The PANAS contains only all high arousal-level emotions that may are not considered emotions (for example, the words *active* and *strong*), while SPANE reflects all levels of arousal for both positive (joy, happy, contented) and negative emotions (sad, angry, and afraid).

d. Positive and Negative Affect Schedule

The Positive and Negative Affect Schedule (PANAS) is made up of two 10-item scales developed by Watson, Clark, and Tellegen (1988). PANAS is a short and easy to administer scale that measures positive emotions (PA) and negative emotions (NA). PANAS was reported to be highly internally consistent. PA and NA were reported uncorrelated, and stable over the 2-month period (Watson et al., 1988). PA and NA are factors that have been identified within individual analysis and used across various experiments of emotional experience (Watson & Clark, 1994). Distress and unpleasantly engaged are features of high NA, while low NA is described by the absence of those emotions. Contrast to NA; PA represents the extent to which individuals are pleasantly engaged with the environment. However, according to (Watson & Clark, 1999), Positive emotions

and negative emotions are better characterised as positive activation and negative activation because both are predominantly defined by the *active* positive and negative emotional states with an absence of low arousal states.



Worksheet 3.1 The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988)

PANAS Questionnaire

This scale consists of a number of words that describe different feelings and emotions. Read each item and then list the number from the scale below next to each word. **Indicate to what extent you feel this way right now, that is, at the present moment *OR* indicate the extent you have felt this way over the past week (circle the instructions you followed when taking this measure)**

1	2	3	4	5
Very Slightly or Not at All	A Little	Moderately	Quite a Bit	Extremely

_____ 1. Interested	_____ 11. Irritable
_____ 2. Distressed	_____ 12. Alert
_____ 3. Excited	_____ 13. Ashamed
_____ 4. Upset	_____ 14. Inspired
_____ 5. Strong	_____ 15. Nervous
_____ 6. Guilty	_____ 16. Determined
_____ 7. Scared	_____ 17. Attentive
_____ 8. Hostile	_____ 18. Jittery
_____ 9. Enthusiastic	_____ 19. Active
_____ 10. Proud	_____ 20. Afraid

Scoring Instructions:

Positive Affect Score: Add the scores on items 1, 3, 5, 9, 10, 12, 14, 16, 17, and 19. Scores can range from 10 – 50, with higher scores representing higher levels of positive affect. Mean Scores: Momentary = 29.7 (*SD* = 7.9); Weekly = 33.3 (*SD* = 7.2)

Negative Affect Score: Add the scores on items 2, 4, 6, 7, 8, 11, 13, 15, 18, and 20. Scores can range from 10 – 50, with lower scores representing lower levels of negative affect. Mean Score: Momentary = 14.8 (*SD* = 5.4); Weekly = 17.4 (*SD* = 6.2)

Figure 1.2. Positive and negative affect schedule (PANAS)

According to Google Scholar, there were approximately 7500 articles in 2016 that used the PANAS scale. In the next section we describe some critiques of PANAS but note that these also apply more generally to the other scales described here, as they are all based on recognition.

Critique of PANAS. Compared with previously described emotional tests, there is a gradual increase in the number of studies using the PANAS scale. The investigation of PANAS measures is now commonly used with different populations, like children (Laurent et al., 1999), parents (Ebesutani, Okamura, Higa-McMillan, & Chorpita, 2011), and adolescents (Huebner & Dew, 1995; Wilson, Gullone, & Moss, 1998). Some research has also modified PANAS for specific populations such as college students (Killgore, 2000), youth sport (Crocker, 1997), and inpatient medical rehabilitation (Ostir, Smith, Smith, & Ottenbacher, 2005). Some research has also explored these findings across dozen of cultures (for example: Merz et al., 2013; Carvalho et al., 2013; Mihić,

Novović, Čolović, & Smederevac, 2014).

Though PANAS has become a commonly used measure of emotions, it is not without criticism, particularly in the way stimuli are presented. As can be seen from previously described emotional scales, including PANAS, their respondents are framed in terms of specific feelings and there are always responses about the unfamiliarity of the emotion terms. Another criticism of PANAS focuses on the emotion terms used. Half of the positive affect of PANAS contains words that are not considered as emotions (like *strong, alert, inspired, determined*, and *active*), while common emotion words (like *happy* and *sad*) are excluded.

1.4. The Importance of Recall Based Emotion Scale

All recognition-based scales, as described above, naturally constrain the range of emotions that can be expressed by the participant. This is described further in the following chapter. However, these scales are also not content neutral. In other words, they assume that the stimuli presented to participants have similar emotions responses and can be evaluated with equivalent accuracy across participants. For example, Grove and Prapavessis' (1992) study of POMS found some participants struggled in accepting terms *bushed* and *blue*. One can always level similar criticisms of other scales that use specific words. And, of course, this may also differ by age, culture, or language. These are natural constraints of recognition-based scales.

One approach to better understanding these constraints is to provide a new emotional scale that is framing-free and allows the participants the ability to express their "true" feelings. This can be accomplished by constructing an emotion scale based on recalled emotions.

Van Rensbergen, Kuppens, Storms, and De Deyne (2015) study is an example of the prior effort to use recall-based scale in assessing individual differences. This experiment asked participants to describe their personality using any 10 adjectives. Participants' personality scores were obtained from the average correspondence of their responses to the Big Five personality factors, using the responses' semantic similarity to words for which trait correspondence

was already known.

Van Rensbergen, Storms, and De Deyne (2015) found that a recall-based approach has the advantage of being a more immediate and natural task since participants are allowed to give any response they consider meaningful. Moreover, when a recall-based scale is used in addition to a recognition-based questionnaire, the validity was significantly raised because relevant self-knowledge is activated by making people think consciously about their personality (Claeys, De Boeck, Van Den Bosch, Biesmans, & Böhrer, 1985). The additional advantage of a recall-based test is that it took considerably less time per participant, compared to responding to a large number of Likert scale items.

Based on the general observation that there appears to be a difference between recall and recognized emotions, past studies have focused exclusively on recognized emotions, and that recall-based assessments of individual differences may generate valid results, this thesis focuses on developing a new method to recover human emotional states based on free recall of recently experienced emotions. This method is called the *Emotional Recall Task*.

1.5. Thesis Outline

This thesis focuses on how people search their emotional memory and how this is related to other measures of emotion based on recognition. The outline for my thesis is organized as follows:

In this chapter, Chapter 1, I have pointed out the value in developing a new method to recover human emotional states based on emotional free association called the Emotional Recall Task (ERT).

In Chapter 2, I will describe the ERT 1.0, explaining how it uses the emotions participants produce and word valence norms to compute a measure of affective state. This is then validated against other well-being scales, showing that the ERT is on par with but differs from PANAS.

Chapter 3 will provide a description of the ERT 2.0, the recently developed version of ERT 1.0. This chapter shows how the ERT 2.0 can avoid valence norms by having participants rate the emotional valence and arousal of the words they produce.

Chapter 4 describes the test–retest reliability of the ERT and also discusses the longevity of specific emotions.

Chapter 5 describes interviews of participants immediately following the ERT. This helps to explain how emotional memory is searched and how the ERT can be used to establish differences in the dimensionality of emotions, further expanding on the search (retrieval) phase of emotional memory search.

Chapter 6 demonstrates how the ERT 2.0 can be used in cross-cultural comparisons using a sample of more than 1000 Indonesian participants. Specifically, this chapter compares Indonesian and American ERTs.

Chapter 7 provides an investigation on the relationship between the emotions we recently experienced (ERT) and the emotions we can recall (emotional literacy). This chapter also evaluates the peak-end rule in relation to the ERT and recent emotional experiences.

Chapter 8 concludes this thesis by discussing some future directions and the newly proposed scale “ERT” that encompasses the critical measurement of emotion retrieval.

Chapter 2

The Emotional Recall Task: Juxtaposing Recall vs. Recognition-based Affect Scales

"How people recall and estimate their emotions is an important component of people's self-concepts and how they conceptualize their lives"
(p. 292, Thomas & Diener, 1990).

New emotions scales often originate when limitations are identified in existing emotions scales (Watson & Clark, 1999; Lucas, Diener, & Larsen, 2003; McDowell, & Praught, 1982; Thompson, 2007). Because all existing emotions scales are recognition based, previously identified limitations have often involved complaints that the list of terms on which participants base their emotional judgements "do not capture the range of people's experienced emotions" (Diener et al., 2009). In other words, the emotions that people experience are not those on the recognition scale. A scale based on recalled emotions might be a better indicator of people's emotions across a broad range of emotions. Moreover, such a scale, by revealing where it is not predictive of people's recognized emotions, would offer insight into how emotions are accessed and the dimensionality of recalled emotion versus recognized emotion. In this chapter, we introduce a recall-based emotion scale, the Emotional Recall Task, and compare it with a number of currently popular recognition-based scales. Before introducing this task, we first briefly discuss the need for a recall-based emotional scale motivated by the history of research on emotional dimensionality. We then explain the potential differences in the memory literature between recall and recognition as they apply to emotions.

If people experience emotional dimensionality in different ways, this potentially throws existing affective measurement scales into question. This is because the most popular approach to measuring emotions is to ask people about their ability to recognize how much they felt each of a set of emotions provided on a pre-determined checklist. Such recognition-based scales make two overarching assumptions. The first is that people will be able to identify their own emotions in relation to the words provided in the checklist. This we call the assumption of emotional specificity. The second is that the checklist will adequately cover a person's experience of emotions. This we call emotional breadth.

To put the ideas of emotional specificity and breadth in context, let us consider what is arguably the most widely used recognition-based checklist, the Positive and Negative Affect Schedule (PANAS) (for review see Diener et al., 2010). The original article describing PANAS (Watson et al., 1988) currently has more than 7500 citations as reported by Google Scholar, which is a measure of PANAS's utility as an emotional scale. Even though PANAS has become a commonly used measure of affect, it is not without criticism. Because emotional stimuli are presented, they may frame respondents in relation to certain emotions which may be more or less familiar to them (e.g., Diener et al., 2009). Another criticism of PANAS focuses on the specific terms, some of which are not generally considered as emotions (*strong*, *alert*, *inspired*, *determined*, and *active*), while common emotion words (*happy* and *sad*) are excluded. Five of the terms in PANAS focus on anxiety, and there are few low arousal terms (Diener et al., 2009).

Though PANAS is only one among many recognition-based scales (see Chapter 1), its potential problems are likely to be common to recognition-based scales more generally, including issues of breadth and specificity in addition to order and priming effects (e.g., Hansen & Schantz, 1995; Wang, Busemeyer, Atmanspacher, & Pothos, 2013). For example, being reminded of a forgotten emotion may make that emotion more salient than it otherwise would be in day-to-day experience.

One way to overcome these problems is to allow individuals to freely recall emotions they have recently experienced (e.g., in the last month). Because emotional terms are highly salient in free recall tasks (Altarriba & Bauer, 2004),

the experience of an emotion may be easily recalled. Moreover, the recollection of emotional memories in a free recall task may be a better indicator of general emotional states and well-being than recognition-based scales because they reflect the emotional pathways laid down in the associative memory network (Bower, 1981), which plays a substantial role in the recollection of experience.

2.1. Development of the ERT Scale

In this investigation, we propose an alternative approach that encourages people to actively search their memory for emotions they have experienced. Participants are required to first produce 10 words to best describe their feelings over a recent period of time. Next, they rate each of these words on a 100-point scale to indicate how frequently they have experienced these emotions.

To retrieve valence information from the words produced in the ERT task, we rely on an extended version (Warriner, Kuperman, & Brysbaert, 2013) of Bradley and Lang's (1999) Affective Norm for English Words (ANEW), which provides ratings of valence for almost 14,000 English words. Each word was rated on a scale from 1 (unpleasant) to 9 (pleasant). This database allows us to transform a list of emotion descriptors collected from each participant into a vector of valence.

The overall emotion state of each participant in the ERT is calculated using the formula below:

(1) $$\mathrm{V} = \frac{1}{10}\sum_{i=1}^{10}(V_i - 5)\times R_i$$

where V denotes overall emotion states of an individual in terms of valence. R represents the reported frequency of the i^{th} emotion. V_i denotes the respective valence rating of i^{th} emotion in the extended ANEW.

Participants generated 139 words that cannot be transformed into valence ratings because those words are not included in the emotion norm database (Warriner et al., 2013). To tackle this issue, we used Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013), a language model, to replace them with the most proximal words that exist in the extended version of ANEW.

2.2. Subjects and Measures

Participants were recruited from Amazon Mechanical Turk. They are based in the United States and reported as native English speakers. We excluded 4 participants from analysis because they failed to follow instructions. This left us with 126 participants (male = 57, female = 69).

The questionnaire was administered using Qualtrics. Following the consent form, participants were taken to a webpage and provided with the following instruction: "Please list 10 words that best describe your emotions in the past month". After completing the entry of these 10 words, a second page appeared with the 10 words participants produced in the first page and an instruction asking participants to "indicate how frequently you have experienced each of these emotions on the slider below". The slider ranged from 0 (not often at all) to 100 (very often). The sequence of emotion terms was randomized instead of following the sequence of recall.

2.3. Recall Pattern in the ERT Scale

Participants produced a wide range of words to describe their emotions. In total, 466 unique words were generated and 64% of them were mentioned only once.

Our analysis shows participants tended to first recall emotions they experienced more frequently (Fig. 2.1A). The first few emotions in the recall sequence are also produced more quickly than the rest: starting from the 5th word, the amount of time spent on searching for words to describe emotions starts to increase (Fig. 2.1B). In addition, the valence of emotion terms is bimodally distributed, suggesting that people experience more emotion-loaded term than neutral ones (Fig. 2.1C).

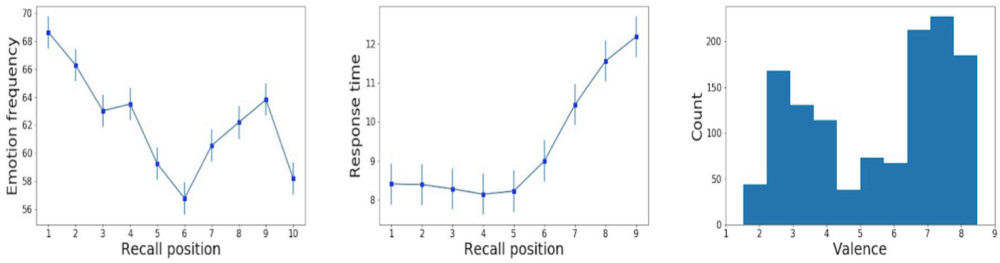


Figure 2.1. From left to right, the averaged frequency of experiencing reported ERT emotions in each recall position; the averaged time (in seconds) spent on generating ERT emotion words in each recall position; distribution of valence of all terms produced during the ERT task

How is the ERT different from the PANAS in terms of emotional breadth and emotional specificity? Figure 2.2A shows that most PANAS terms are not among the most popular words that people frequently use when describing their past emotions. Only 1 PANAS term (*excited*) appeared among the top 10 most frequently recalled emotions. This raises concerns that participants may not have identified their past emotions in relation to many of the terms in the PANAS. Furthermore, consistently with previous criticism on the PANAS, our result presents quantitative evidence that the PANAS suffers from issues of emotion breadth. Figure 2.2B compares distribution of the PANAS terms and the ERT terms on the emotion space of valence and arousal. It shows that ERT terms distribute across the entire arousal space while the PANAS contains no low arousal emotion term. Moreover, although both scales cover two extreme ends of valence space, the PANAS does not seem to have enough neutral terms that might be important to the well-being of many people.

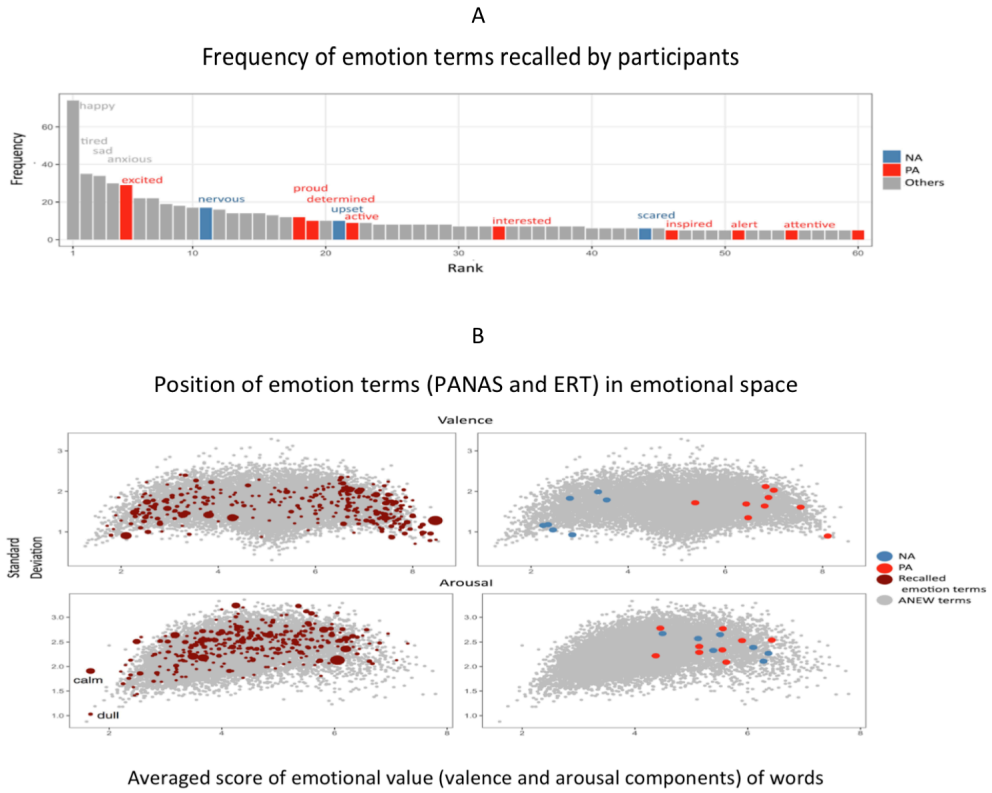


Figure 2.2. Emotional breath and specificity of the ERT and the PANAS

Figure 2.2A shows the frequency of words recalled in the ERT and where the PANAS words are located in the ERT frequency ranking (highlighted in red and blue respectively for positive and negative emotion). Figure 2.2B shows where the PANAS terms and the ERT terms are located along the dimensions of valence and arousal. The x-axis is the mean emotional rating and the y-axis is the standard division of these ratings. Higher standard deviation indicates larger degree of disagreement in the emotional connotation of words. Each grey dot represents one word from the existing emotional norm database (Warriner et al., 2013) and together they define the emotional space for English words.

Another difference between the ERT and PANAS is how they aggregate emotion scores. The PANAS asks participants to report "the extent you have felt in this way" on a scale from 1-5 and then sums up the scores for positive and negative emotion. This approach assumes that experience of different types of emotion would have the same weight on one's overall emotional state. However, this assumption is not likely to hold because there are always some emotions that have greater impact than others. For example, as suggested by the average valence rating of *nervous* and *scared*, *nervous* may have less negative

psychological impact than *scared*. In contrast, the ERT solves this issue by assigning each emotion term a valence value that quantifies how pleasant this term is based on the Warriner et al.'s (2013) database: *nervous* is assigned to a valence value of 3.56 while *scared* received a more negative valence of 2.79.

2.4. Convergent Validity

Convergent validity can be established by compare the ERT and the PANAS in terms of their correlation with measures of related constructs, such as well-being, depression and anxiety. We have used the ERT scale in conjunction with the following commonly used measures, as described below.

The Positive and Negative Affect Schedule (PANAS, Watson et al., 1988). PANAS consists of two 10-item scales. It was developed to provide a brief measure of positive and negative emotion. The 20 PANAS items were derived from a principal component analysis of Zevon and Tellegen's (1982) 60-item emotion checklist. Respondents are asked to rate the extent they experienced each emotion within a specific time frame, with reference to a 5-point scale that ranges from 'very slightly or not at all' to 'very much'. Different time frames (e.g., "right now", "today", "during the past few days", "during the past week", "during the past few weeks", "during the past year", "in general") have been used with the PANAS. In the present study we set time frame to "during the past month". The PANAS scale intercorrelations and internal consistency reliabilities (Cronbach's coefficient α*)* are all acceptably high, ranging from .86 to .90 for positive emotion and from .84 to .87 for negative emotion (Watson et al., 1988). The reliability of the scales is clearly unaffected by the time instructions used.

The Ryff Scales of Psychological Well-Being (SPWB, Ryff & Keyes, 1995). The SPWB is a theoretically grounded instrument that specifically focuses on measuring multiple facets of psychological well-being. These facets include the following: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. Individuals respond to various statements and indicate on a 6-point Likert scale on how true each statement is of them. Higher scores on each scale indicate greater well-being on that dimension. We used the 18-item version in the current study.

The Diener Satisfaction with Life Scale (SWLS, Diener, Emmons, Larsen, & Griffin, 1985). The SWLS is a short 5-item instrument designed to measure global cognitive judgments of satisfaction with one's life as a whole. The scale does not assess satisfaction with life domains such as health or finances but allows subjects to integrate and weight these domains in whatever way they choose.

UK Office for National Statistics Well-Being Measurement (ONS, Tabor & Stockley, 2018, May 17). The ONS was developed by the Office for National Statistics of UK to assess personal well-being using 4 measures that capture 3 types of well-being: evaluative, eudemonic and experience. These measures ask people to evaluate the overall life satisfaction, worthiness of things they do, happiness, and anxiety. It was first added to the Annual Population Survey (APS) in April 2011 and has been used in many surveys across the UK.

The Depression Anxiety and Stress Scale (DASS-21, Lovibond & Lovibond, 1995). DASS consists of three 7-item self-report scales that measure depression, anxiety, and stress correspondingly. Each item was rated on a 4-point scale.

The Beck Depression Inventory (BDI, Beck, Steer, & Brown, 1996). The BDI measures severity of depression in normal and psychiatric populations. The questionnaire was developed from clinical observations of attitudes and symptoms occurring frequently among depressed psychiatric patients and infrequently in non-depressed psychiatric patients. The questionnaire contains 21 questions, each ranging on a scale from 0 to 3.

2.5. Correlation between the ERT and the PANAS

A good affect scale should be able to predict related constructs. We first analysed the relation between the ERT and PANAS. The pairwise correlation coefficient of PA, NA and the ERT can be found in Table 1. Consistent with previous studies of PANAS, we found the PA and NA component are independent of each other (r=-.14, p>.05). The ERT correlate with both PA and NA at similar strength (r=.56, p<.001 for PA and r=-.59, p<.001 for NA). This is expected

since the ERT is designed to be able to capture the entire spectrum of emotion space.

Table 2.1. Correlation table between all measures

	ERT	Panas		SWLS	PWB	ONS				BDI	DASS		
		PA	NA			LifeSatisfy	LifeWorthy	Happy	Anxiety		Depression	Anxiety	Stress
ERT													
Panas PA	0.57***												
Panas NA	-0.58***	-0.14											
SWLS	0.65***	0.57***	-0.26**										
PWB	0.58***	0.43***	-0.57***	0.54***									
ONS_LifeSatisfy	0.75***	0.64***	-0.43***	0.87***	0.61***								
ONS_LifeWorthy	0.69***	0.61***	-0.39***	0.72***	0.68***	0.83***							
ONS_Happy	0.70***	0.53***	-0.47***	0.63***	0.56***	0.78***	0.71***						
ONS_Anxiety	-0.52***	-0.17	-0.73***	-0.25**	-0.53***	-0.39***	-0.36***	-0.52***					
BDI	-0.69***	-0.36***	0.75***	-0.44***	-0.64***	-0.58***	-0.55***	-0.58***	0.67***				
DASS_Depression	-0.69***	-0.38***	0.68***	-0.45***	-0.69***	-0.62***	-0.66***	-0.65***	0.63***	0.87***			
DASS_Anxiety	-0.39***	-0.03	0.65***	-0.05	-0.47***	-0.2*	-0.24**	-0.27**	0.58***	0.64***	0.72***		
DASS_Stress	-0.53***	-0.21*	0.71***	-0.29***	-0.54***	-0.47***	-0.48***	-0.53***	0.63***	0.75***	0.83***	0.81***	

Note: *p<.05,**p<.01,***p<.001

We further explored the discrepancy between the ERT and the PANAS by analysing participants whose emotional states are inconsistent between the two measures. Figure 2.3A shows how participants' ERT scores are related to the PANAS. For example, in the ERT task, participant 15 (ID number = 15) generated a number of negative emotion terms and no positive terms and reported experiencing each of the negative terms with high frequency (Fig. 2.3B2). Yet this participant reported extremely low negative affect in the PANAS scale (Fig. 2.3A left). Similarly, participant 72 recalled 8 positive emotions, 1 neutral emotion and 1 negative emotion (Fig. 2.3B4). But the same participant's PANAS score suggests he experienced little positive affect. Examples like these suggest that positive and negative affect may be underestimated in the PANAS because participants' experienced emotions are not on PANAS's emotion list. See similar results for participants 75 and 66 (Fig. 2.3B1 and B3).

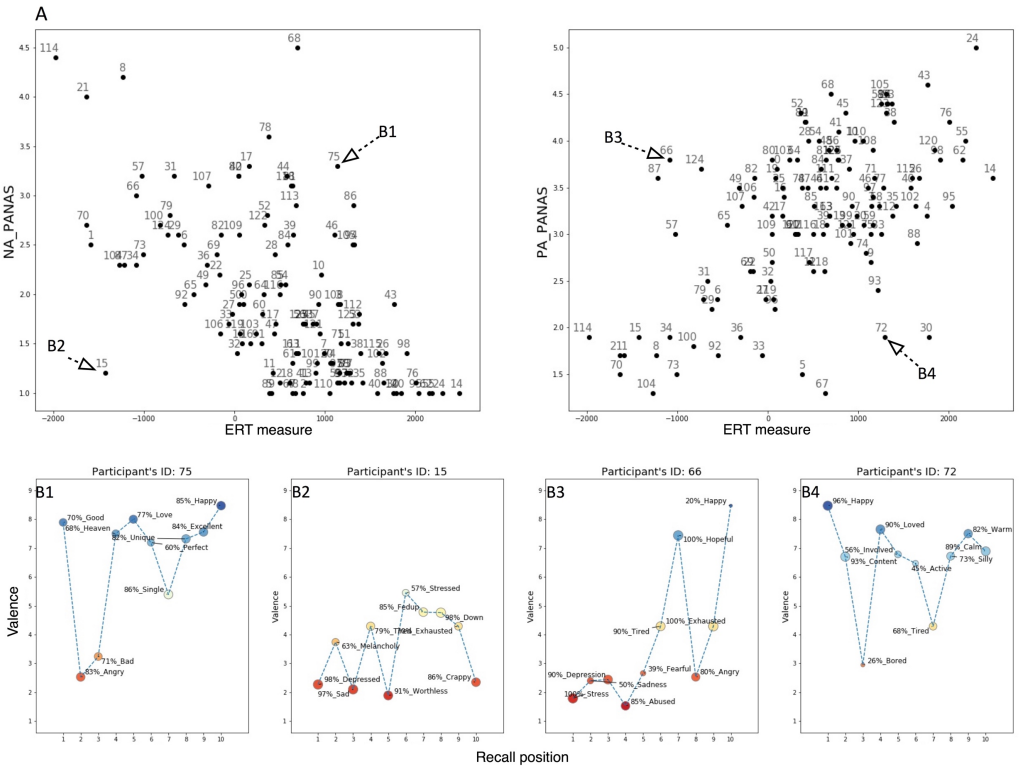


Figure 2.3. Discrepancy between the ERT measure of emotion and the PANAS

Figure 2.3A shows correlation between ERT measures and NA and PA of the PANAS. Figure 2.3B1-2.3B4 shows the sequence of 10 words produced by the 4 participants identified in A and also provides their frequency (in %) next to each entry. Colour shows word valence (blue = positive, red = negative) and dot size corresponds to frequency.

2.6. Correlation with other related constructs

To test validity of the ERT, we compared it with the PANAS on how well they predict related constructs. The correlation matrix is shown in Table 1. The ERT performs at least as well as the PANAS in indicating the wellbeing-related constructs (Diener, Ryff and ONS4), and 2 depression measures (BDI and DASS), while the NA of PANAS performs better in indicating anxiety and stress. This is not surprising since 4 out of 10 terms in the NA of the PANAS is anxiety related (Diener et al., 2009).

Next, to test whether 10 words is sufficient to capture emotion experience in the ERT, we performed a sensitivity analysis to show how correlational

strength between the ERT measure and other constructs change in relation to the number of emotion terms included. Figure 2.4 shows that the correlation improves when including more subsequently recalled emotion words. This improvement has a diminishing marginal return: the improvement plateaus after inclusion of approximately the first 7 words, depending on the scale compared.

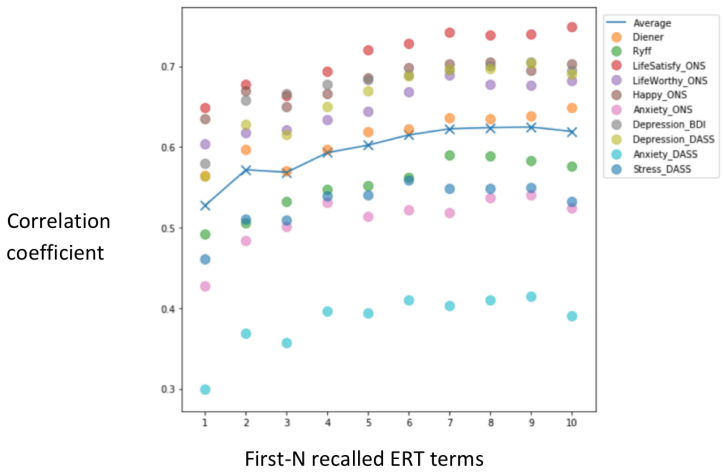


Figure 2.4. Sensitivity analysis for correlation coefficient between the ERT measure and other constructs in relation to increasing number of the ERT words included (in recall order)

2.7. Discussion

We found that the Emotional Recall Task, by relying on recalled memory of emotion experience, effectively captures emotional states and correlates highly with other commonly used measures of well-being. The ERT performs as well as PANAS in predicting 3 different measures of well-beings (Diener, Ryff, ONS) and two measures of depression (BDI and DASS). PANAS outperforms ERT in predicting anxiety and stress. In addition, both correlation (between PANAS and ERT) and individual case studies (Figure 2.3) shows that ERT captures different aspects of emotional experience from the PANAS. This could be due to a lack of emotional specificity and breadth in recognition-based emotion scales.

One important limitation to PANAS and other existing recognition-based affect scales is their limited generalisability. People across various cultural and social groups often have systematic differences in their experienced emotion

space. For example, *schadenfreude* is a German word (means derived happiness from other's pain) for which there is no such word in English. Scollon, Diener, Oishi, and Biswas-Diener (2004) have identified emotion terms important and specific in certain cultures and that do not exist in other cultures. Even within the same culture, the PANAS can be problematic when comparing scores across different social or age groups. Because PANAS lacks low-arousal terms, young people may score higher on positive emotion than the elderly just because they are generally more sensation-seeking (Oishi, Schimmack, & Colcombe, 2003). The elderly may score higher on pleasant terms such as 'contented' and 'peaceful', but these low-arousal feelings are not assessed by the PANAS. Therefore, a checklist approach to emotion scale can hardly be applied to all cultures or even all groups within a same culture because people vary from one to another in terms of what emotions were experienced and valued.

In contrast, the ERT overcomes this problem by allowing all individuals to freely choose emotion terms that best summarise important experiences in their life. It functions like a Swiss army knife that adapt the scales accordingly to best fit each individual's emotion experience without losing any affective (valence) information. At the macro level, the ERT scores can be used to make comparison across individuals just like other recognition-based scales, while at the micro level, the ERT offers greater details on the entire emotional space one has experienced instead of a pre-determined and often insufficiently broad emotion space.

Based on the finding that conceptual processes influence how sensory information is sampled from the physical surroundings (Sowden & Schyns, 2006). It is possible that a predetermined emotion checklist would influence how people sample/recall past memories to construct an emotional percept. Based on the claims that emotion words (with associated concepts) that become accessible reduce ambiguity inherent in most facial behaviours and facilitate quick and easy perception of emotions (Barrett, Linduist, & Gendron, 2007), it is possible that presence of emotion words may interact with the process of evaluating past feelings. The ERT, by removing a predetermined emotion checklist, avoids activating emotion-loaded concepts. This is perhaps why some participants (e.g. participant ID 15 in Fig 2.3B1) recalled a lot of positive emotions and few

negative ones in the ERT but reported intense experience of negative emotion in the PANAS scale.

In closing, the ERT is a simple and valid way to measure emotion. Relying on recall process, it caters to individuals' specific emotional experience and therefore avoids several disadvantages shared by all recognition-based scales, namely, emotion specificity, emotion breadth, and priming. In addition, in the current investigation, the richness of the information embedded in the responses (such as the order of words produced, time interval between two productions) in the ERT are all reduced to one single dimension: valence. Perhaps that is the reason why it performs well in predicting well-being related constructs but not so accurate in predicting a specific emotion experience such as anxiety and stress.

Finally, the ERT confirms a central question of this thesis: A recall-based affect scale can capture the central tendencies of recognition-based emotion scales. Moreover, the ERT appears to sit reasonably well within the various scales administered, indicating that its content neutrality (not focusing on any specific component of emotion, such as life satisfaction or negative emotion) allows it to capture a more general measurement of emotional state.

Chapter 3

The Emotional Recall Task 2.0: Self-Rated Version

3.1. Background

Emotional scales can be classified as recall-based or recognition-based scales. The standard format is typified by the positive and negative schedule (PANAS: Watson et al., 1988) and the scale of positive and negative emotional experience (SPANE: Diener et al., 2010). These are large usage examples of recognition-based emotional scales that require participants to recognize emotional words and evaluate the frequency that they experienced them in the recent past. The Emotional Recall Task, in comparison, is a recall-based scale that allows participants to provide the first emotions that come to mind in their recent experience. In previous chapter (Chapter 2), we found that recognition-based scales were often less well-correlated with one another than they were with the ERT. This suggests that the ERT may be a more accurate and general test of emotion.

The previously investigated version of the ERT (here called ERT 1.0 and reported in Chapter 2 of this thesis) used emotion ratings from publicly available word norms (Warriner et al., 2013). This has two problems. First, the word valences used to compute the ERT 1.0 score are not specific to individuals producing the words. Second, when researchers use the ERT 1.0 they can only use participants' data when the words the participants produce can be matched to words in the emotion rating norms (Warriner et al., 2013).

One potential solution to these problems is to have participants rate their own words. This allows researchers to use all the words participants produce and to know the participants rating of the associated emotional state. Effectively this turns the ERT 1.0 into a self-rated task, with results that can be calculated almost instantly following assessment and without the need of any sophisticated computations or additional data such as word norms.

This chapter describes this version of the ERT (ERT 2.0) and show that it outperforms the original ERT 1.0. Before doing that, I will first describe the ERT research up to this point.

Emotional ratings are popular in emotional research as they reveal individual differences that may influence people's performance in a variety of contexts. Emotional ratings ask participants to rate words, like *love* and *eager* on a Likert-type scale based on how negative or positive the word is. These scales exist for numerous languages. In English, there are scales with up to 43,000 words (e.g., Warriner et al., 2013; see also Bradley & Lang, 1999). These scales are widely used to understand how emotional content of words influences word processing and memory (Kensinger & Corkin, 2003a, 2003b), how emotions are produced and perceived (Scherer, 2005), how emotional features influence word and memory processing (Bock & Klinger, 1986), and how texts signal the emotional state or sentiment of their authors (Pak & Paroubek, 2010, May).

The application of rating scales used to derive the attitude towards particular objects, events or concepts was pioneered by Osgood's (1952) semantic differential scale. This scale required subjects to rate verbal stimuli on 50 opposite bipolar scales (e.g., hot-cold, white-black, fast-slow). More recently, one of the most cited affective rating norms in Google Scholar 2018 is the Bradley-Lang (1999) scale, which developed a set of verbal materials that have been rated in terms of pleasure, arousal, and dominance in the English language. This is called the Affective Norms for English Words (ANEW). The development of ANEW complemented the existing IAPS (International Affective Picture System; Lang, Bradley, & Cuthbert, 1997) and IADS (International Affective Digitized Sounds; Bradley & Lang, 1999), which are collections of picture and sound stimuli, respectively, that also include these emotion ratings for about 1000 words.

Most of the research above developed from the dimensional theoretical perspective that describes humans' emotional experience in continuous terms such as variations in valence and arousal. Valence indicates the hedonic value of emotion from unpleasant to pleasant, whereas arousal indicates the degree of activation of emotion from calming to exciting (Bradley & Lang, 2000; Russell, 2003).

Using emotion rating scales, we were interested in determining whether or not the emotions people recall could be a valuable indicator of emotional states alongside more standard recognition-based scales. All existing emotional measurement scales use emotion recognition to measure emotional states (e.g., PANAS; Watson et al., 1988; Diener et al., 2010). But these scales often do not list well-known emotional words. For example, PANAS does not list *happy* or *sad* as a response.

In contrast to asking people their ratings based on a pre-determined emotion checklist, we developed a different approach to capture emotions by allowing participants to actively search their emotional memory and produce the first 10 words describing their recent emotions in a task called The Emotional Recall Test (ERT).

In Chapter 2, we found that the ERT outperforms PANAS in its prediction on constructs like well-being, depression, and stress. However, because this first study used emotion values of words based on the Warriner et al. (2013) norms, some of the emotional responses needed to be adapted or removed because the particular emotion word used was not listed in the norms. For example, if a person produced the misspelled word *happie,* this word would need to be changed to *happy* (corrected for misspelling) or removed from the scale in order to compute the participant's ERT score.

There are, however, two additional issues. One issue is when individuals produce words that are not in existing norms like *equanimous* or *wife* (both observed in previous participants' data). These words represent problems because we are unlikely to accurately infer the underlying emotional value of these words (is *wife* positive or negative, and by how much?). They also point to another problem, which is that people may mean different things with words even when they are on existing scales. Some individuals may experience *equanimous* as verging on nirvana, whereas others may experience it as near

boredom, and others may not know understand the meaning accurately. A second issue is that the existing ERT requires that researchers have access to appropriate valence norms, which may not always be the case for cross-cultural or developmental studies.

To solve these problems, we propose a method for capturing emotional ratings for words by simply asking participants to rate the recently experienced emotions they produce. In this modified version of the ERT, participants produce ten emotions they have felt in the past month, and then they rate these emotions for how often they have felt them, and finally they rate each emotion for its valence (and additionally for arousal) on a scale from 1 to 10.

3.2. Approach/Methodology

a. Instruments explanation

The Emotional Recall Test (ERT) 1.0. The ERT 1.0 contains two parts. In the first part, participants were instructed to write ten distinctive words describing their experienced emotions during the last month. In the second part, the participant is instructed to rate "how often" they experienced each of the represented emotions on a scale from *not often* (0) to *very often* (100).

The Emotional Recall Task (ERT) 2.0. Unlike the ERT 1.0, where we transformed participants' emotion into ANEW ratings on valence and arousal, participants in the ERT 2.0 are asked to provide the valence and arousal rating for each of their experienced emotions. This adds two additional sections to the ERT 1.0. Section 1: participants are asked to provide 10 distinctive words describing their *emotion*s in the past month. Section 2: the 10 words produced in Section 1 are presented in a random order with the instructions to rate how often have they experienced each of these *emotion*s. Participants are then asked to give their ratings on valence (Section 3) and arousal (Section 4) by using a slider on a scale from 1 to 10. For the valence instruction, participants were asked to rate the pleasantness or unpleasantness of their emotion. For the arousal, participants were instructed to rate how calm or intense was the emotion. The arousal rating is not intended for future use with the ERT 2.0, but we included it in this first

study in order to confirm that the words participants produced had an emotional breadth broader than that reported for PANAS, which is what we found in Chapter 2.

The Positive and Negative Affect Schedule (PANAS). The PANAS (Watson et al., 1988) consists of 20 items rated on a 5-point Likert scale of 1 (not at all) to 5 (extremely). The scores are summed across items on the positive-negative scale to produce a score between 10 and 50.

Other Validation Scale. To establish the validation of ERT 2.0, a correlation analysis is calculated with other variables like 5-item subjective well-being instrument (The Diener Satisfaction with Life Scale), 4-item personal well-being instrument (The UK Office for National Statistics Well-Being Measurement), and 21-item depression-anxiety instrument (The Beck Depression Inventory).

b. Procedure

The ERT 2.0 was built using the Qualtrics online survey platform. Experiments begin with the information sheet and the participants' consent statement. Following this, participants were presented with The ERT 2.0, PANAS, Diener's Satisfaction with Life Scale, ONS, BDI-21, and demographic questions. ERT is always presented first so as not to prime participants with emotion words from the other scales. Demographic questions included first language, country of origin, ethnicity, gender, age, marital status, and education. Participants completed the experiment within half an hour.

c. Participants Demographics

The ERT 2.0 were presented to 200 workers of Amazon Mechanical Turk (MTurk). The MTurk platform was chosen in order to make sure that all the participants selected are first language English speakers. Each participant was compensated with $3.00 according to MTurk payment regulations. 27 participants were removed due to error in the data and improperly submitting their responses. 13 participants were removed because they indicated their

primary language was not English (this was selected in MTurk and so was a predefined selection criteria). The final data consisted of 160 participants. Participants reported as female in 57 cases (35.4%) and as male in 103 cases (63.4%). The mean age of participants was 32 years (*M*=32.4; *SD*=9.7) ranging from 18 to 69 years.

3.3. Results

a. ERT Scores

Participants' ERT score is a weighted sum of the valence combined with the frequency of experience. The result can be written

(1) $$ERT\ 2.0\ SCORE = \frac{1}{n}\sum_{i}^{n=10}(V_i - 5.5)\ x\ W_i$$

There are 10 emotion words, each with valence V_i weighted according to the reported frequency, W_i, from Section 2 of the ERT 2.0. Table 2 presents the descriptive statistics of the ERT scores. This is an intuitive scale with a minimum of -450 and a maximum of 450, with higher scores meaning higher average emotion.

Table 3.1. Descriptive statistics of ERT 2.0 scores

Mean	91.36
Median	76.45
Standard deviation	162.00
Standard error of the mean	12.92
Minimum	-368.10
Maximum	450

b. Comparison of ERT 1.0 and ERT 2.0

Figure 3.1 presents the scores for ERT 1.0 and ERT 2.0. The correlation between ERT 1.0 and ERT 2.0 is r=.76, p<.001 for valence and r=.01, p=.90 for arousal. The relationship between valence for the two scales suggests that valence from the ERT self-rated version is represented well by the valence norms. This helps

validate the previous findings for the ERT 1.0 and is also a promising result for the use of ERT 2.0. It suggests that the previous results for the ERT 1.0 in relation to other scales may carry over to the ERT 2.0. Notice also that the range of scale values for the ERT 2.0 is larger than the range for the ERT 1.0. This is consistent with our hypothesis that that ERT 2.0 will better capture participants' true emotions. For example, the lowest scoring individual on the ERT 2.0 provides ten words all of low valence in the Warriner norms (e.g., crappy = 2.35, lifeless = 2.24, angry = 2.53). The individual, however, gives these words even lower valence rating, indicating that they are all of value 1.

Figure 3.1 also shows that participant arousal ratings are not well represented by the norms. This justifies our need to collect this data to confirm the emotional breadth of participants which we previously reported in Chapter 2 for the normed arousal ratings. Figure 3.2 shows a histogram of the arousal scores. Comparing to the rating value in the ERT 2.0, the normed arousal value in the ERT 1.0 is missing the higher and lower arousal part from the participants' responses. This shows that participants experience a wide breadth of arousal across emotions and once again shows that PANAS may not be detecting emotions associated with moderate arousal words that people claim to experience somewhat frequently. This also confirms what we observed in Chapter 2 for the ERT 1.0.

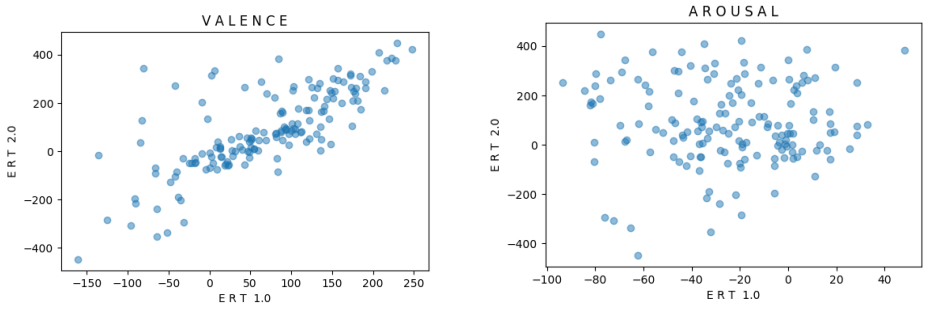


Figure 3.1. Correlations between ERT 1.0 and ERT 2.0 for valence and arousal

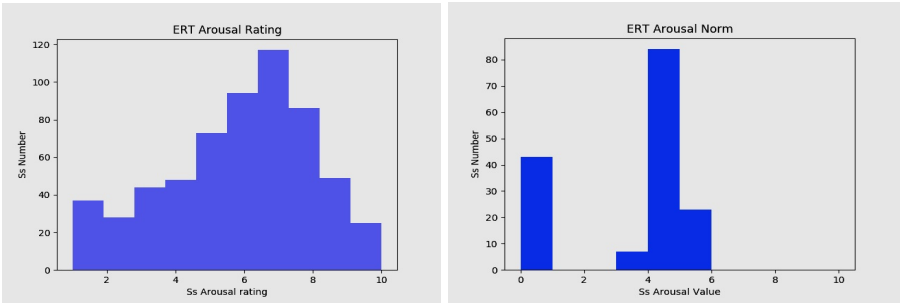


Figure 3.2. Participants' arousal from rating and norm

Table 3.2. Participants with the lowest ERT score

Lowest arousal score in ERT 2.0 (Self-rated): -334.60										
Word	happy	excited	refreshed	relaxed	calm	tired	organized	in love	romantic	anxious
Valence	5	5	5	5	4	5	6	9	6	5
Lowest arousal score in ERT 1.0 (Computer-based): -93.48										
Word	i don't feeling so very happy	respect	punchuality	planning	timing	hardwork	unity	belive	goodthink	helpful
Valence	nan	4.79	nan	3.86	3.4	3.41	4.8	3.33	nan	3.32

c. Frequently reported emotions

The most frequently reported words are shown in table 3.3. Participants reported experiencing more positive emotions than negative emotions. The result also revealed that 343 out of 1570 reported emotions do not have a valence value in the Warriner et al. (2013)'s norms. Some examples of unlisted emotion are *accepted, connected, aspirational, forced, surprised, unmotivated*, and *scattered*. In addition, with 416 words that appear only once, the ERT 2.0 contains 623 unique words.

d. Comparison with PANAS

Among the top ten words produced in the ERT 2.0, only one PANAS word (*excited*) was produced. This word was produced by 19% of participants.

Table 3.3. Frequently reported emotions in the ERT 2.0

Emotion	Frequency	Proportion
Happy	93	0.58
Sad	46	0.28
Tired	34	0.21
Excited*	30	0.19
Anxious	26	0.16
Calm	26	0.16
Hopeful	25	0.16
Worried	20	0.13
Good	20	0.13
Angry	20	0.13

Note: *indicates word from PANAS

Table 3.4 presents the frequency of PANAS words in participants' reported emotions. The table also shown that three of PANAS negative emotions were never mentioned in the participants' ERT 2.0 productions. These words are *distressed*, *ashamed*, and *hostile.*

Table 3.4. PANAS words and proportion of times reported in the ERT 2.0

Positive Affect	*Frequency*	*Proportion*	*Negative Affect*	*Frequency*	*Proportion*
Excited	30.00	0.19	Nervous	10.00	0.06
Determined	17.00	0.11	Afraid	6.00	0.04
Interested	11.00	0.07	Scared	3.00	0.02
Proud	9.00	0.06	Guilty	2.00	0.01
Strong	7.00	0.04	Irritable	2.00	0.01
Alert	6.00	0.04	Jittery	2.00	0.01
Active	5.00	0.03	Upset	2.00	0.01
Attentive	4.00	0.03	Distressed	0.00	0.00
Enthusiastic	3.00	0.02	Ashamed	0.00	0.00
Inspired	1.00	0.01	Hostile	0.00	0.00

e. Does self-rated ERT outperform the norm-based ERT?

The first two columns of table 3.5 show the correlations between ERT 1.0 and ERT 2.0 with each of the other scales. In all cases, ERT 2.0 shows a higher

correlation than ERT 1.0, indicating that ERT 2.0 better captures the ratings from these additional scales.

Overall, ERT 2.0 shows the strongest correlation for PA and the lowest correlation for NA for both valence and arousal calculations. This is contrary to that found for the ERT 1.0, as the valence calculation in 1.0 shows the strongest correlation for PA while the arousal calculation shows the strongest correlation for NA. Finally, we can compare ERT 2.0 with the correlations for each of the other scales. Across all positive emotional constructs, ERT 2.0 shows the highest correlation with the other scales, except for the anxiety subscales of the ONS and the BDI and when scales are well-correlated with their own subscales, such as the ONS subscales.

For the ONS Anxiety and BDI, PANAS NA is the strongest correlation followed by ERT 1.0 on valence. The NA rating (from PANAS) was the most well-correlated for negative emotional constructs, as was found in Chapter 2.

Table 3.5. The correlation matrix between ERT and comparable affect scales

		ERT 1.0		ERT 2.0		Panas			ONS				
		Valence	Arousal	Valence	Arousal	PA	NA	SWLS	LifeSatisfy	LifeWorthy	Happy	Anxiety	BDI
ERT 1.0	Valence	1.00											
	Arousal	-0.08	1.00										
ERT 2.0	Valence	0.76***	-0.02	1.00									
	Arousal	-0.07	0.01	0.14	1.00								
Panas	PA	0.53***	-0.03	0.66***	0.31***	1.00							
	NA	-0.25**	-0.02	-0.08	0.57***	0.14	1.00						
SWLS		0.55***	-0.03	0.65***	0.23**	0.63***	0.26***	1.00					
ONS	LifeSatisfy	0.56***	0.01	0.71***	0.2*	0.61***	-0.01	0.78***	1.00				
	LifeWorthy	0.58***	0.03	0.72***	0.19*	0.64***	-0.12	0.63***	0.81***	1.00			
	Happy	0.64***	0.01	0.73***	0.14	0.59***	-0.08	0.63***	0.77***	0.78***	1.00		
	Anxiety	-0.33***	-0.01	-0.13	0.57***	0.06	0.66***	0.13	0.04***	-0.04	-0.07	1.00	
BDI		-0.38***	-0.06	-0.26***	0.36***	-0.12	0.71***	-0.02	-0.28***	-0.33***	-0.32***	0.54***	1.00

Notes: **p<.05,**p<.01,***p<.001*

3.4. Conclusions

Many scales exist in short and long versions (for instance: the brief version of Ryff's Psychological Well-Being, the Beck Depression Inventory-short version, the Warwick Edinburgh Mental Well-Being Short Form, and the Depression Anxiety Stress Scale short version). This is useful for researchers who may need to estimate constructs but only have limited time with their participants. The ERT 2.0 is aimed to provide researchers with a simplified form for eliciting emotional states from participants, one that does not require the use of valence norms. Though the ERT 2.0 is not shorter (it asks for an additional 10 ratings), it nonetheless provides a more accurate measure of affective state as evaluated with

respect to other existing scales and requires no secondary norm data or restriction to any given language.

The ERT 2.0 also outperforms all scales in relation to its ability to predict positive emotions on other scales (its cross-scale correlation). Negative emotions on the X and Y scales, however, appear to be best captured by the NA component of PANAS. As suggested in Chapter 2, this suggests an important difference between recalled emotions and recognized emotions which may hint at even better scales in the future.

A clear advantage of ERT 2.0 is that participants can produce any words they desire, and researchers can compute ERT 2.0 scores immediately even if they do not have access to a computer. This is potentially valuable for field workers, but it is also promising for using the ERT 2.0 with individuals who do not speak English (see Chapter 6 on ERT Cultural Project).

In sum, the ERT 2.0 offers a novel emotional rating scale based on emotions that appears to capture many aspects of existing recognition-based scales and which also shows a further enhancement over the ERT 1.0.

Chapter 4

Test-retest Reliability of the Emotional Recall Task

4.1 Background

The Emotional Recall Task (ERT) was developed to overcome the drawbacks of existing emotional scales. By allowing participants to freely recall recently experienced emotions, the ERT examines an aspect of emotional memory based on recall instead of recognition. In the Chapter 2 of this thesis, the ERT was found to capture broader emotional breadth and more fine-grained emotional specificity than other scales, such as the PANAS. To provide a more comprehensive psychometric evaluation of the ERT, in this chapter we will investigate the ERT's test-retest reliability.

Test–retest reliability assure that the tool measures the outcome the same way, each time it is used. In this method of reliability assessment, the same scale is applied a second time to the same subjects under conditions as similar as the investigator can make them. The two sets of scores then are correlated, and the index is interpreted as the stability performance of the measures over time (Peter, 1979).

There is one primary assumption in test–retest reliability. This is that the true score does not change between administrations. However, the time interval between the two test administrations is likely to affect test-retest reliability measurements. Over a brief interval, participants may recall information from the first test (carry-over effect) which biases results to be similar based on memory instead of state persistence. Over a long time interval, participants are likely to have changed their state (Allen & Yen, 1979).

To avoid potential retest problems like carry-over effects (i.e., participants give the same emotional words because they remember their initial responses), the length of time between the tests should be carefully arranged (Guttman, 1945). Theoretically, the longer the time-gap the lower the reliability (Heise & Bohrnstedt, 1970), because of the possibility of changes in the population taking the test. When a change in state occurs between the administrations, it is difficult to distinguish change from unreliability (Heise, 1969).

Precise test-retest reliability scores are not calculable because all instruments respond with some error (Bruton, Conway, & Holgate, 2000). Every observed score (O) can be assumed to have a true score (T) and an error component (E) [O = T ± E]. The true score refers to the measurement of one's actual state; while the observed score refers to the score provided by the tool. As it is impossible to know the true score with certainty, reliability can be derived from the ratio of the variance of true score to the variance of observe score. The difference between the true score and the observed score is measurement error. In statistical terms, error refers to all sources of variability that cannot be explained by the independent variable. This ratio, the reliability coefficient, is close to zero when the error component is large, but it is close to one when the error is small. There is no minimum acceptable level of reliability that can be applied to all measures (Bruton et al., 2000).

Additional challenges with test-retest reliability measures are that the test-retest procedure requires data collection from at least two time points. This leads to a higher cost due to multiple administrations of the test and is sometimes impractical as the test population would need to be willing to repeat the test.

Despite these difficulties, test-retest is a valuable method for evaluating the utility of new measurement scales. Moreover, in the case of the ERT, there are existing emotional scales that we can use as benchmarks against which to evaluate the ERTs reliability. That will be our approach here.

Test-retest reliability of emotions may also be influenced by the duration of different emotions. Some emotions have been found to persist for a long time (e.g., days for *sadness*) whereas others tend to fade away quickly (e.g., minutes for *disgust*). Verduyn and Lavrijsen (2015) conducted an experiment to explain whether some emotions last longer than others. In the first section, participants

were asked to fill out a questionnaire on their experienced emotions. To make sure that participants reported emotions and not moods, participants were told that emotions are always elicited by a certain internal or external event that has a clear onset point. Participants were asked to rate the duration and intensity of emotions after recollecting recent emotional episodes. Intensity was rated from 0 (not intense at all) to 6 (very intense). For duration, participants were asked to specify the number of days, hours, minutes, and/or seconds the emotional experience had lasted. Out of 27 emotions, *sadness* lasted the longest, whereas *shame, surprise, fear, disgust, boredom, being touched, irritation,* and *relief* were the shortest emotions. Emotions that are rather short were found to be typically elicited by relatively low importance events (as self-reported by the participant) whereas longer lasting emotions tended to be elicited by important events (Verduyn & Lavrijsen, 2015). Scherer and Wallbott (1994) investigated 7 emotions (*joy, anger, fear, sadness, disgust, shame*, and *guilt*) and found that the emotions significantly differed in duration, with a rank order of duration being *fear = disgust = shame <= anger < guilt < joy < sadness*. Again, this suggests that sadness is the longest lasting emotion.

Previous tests have investigated general test-retest reliability. Correlation for the test-retest reliability for PANAS was .58, $p<.05$ for positive emotion (PA) and .48, $p<.05$ for negative emotion (NA) (Watson et al., 1988). This study had 101 participants. They filled out the PANAS scales with the temporal instruction being emotions felt over the past year, past few weeks, past few days, today, general, moment, and weeks. Following a 1-week break, the PANAS scales were re-administered in the same sequence. Result showed that in the brief time interval (e.g., moment and today) the scale is sensitive to mood fluctuations, however, in the longer time interval (e.g., past year and general) the scale exhibit trait-like stability. This of course may be more a property of what PANAS measures (a dispositional component of emotion) than the persistence of people's 'true' emotional states.

Various studies have been conducted on the stability of the SWLS. Diener et al. (1985) reported a correlation of .82 for test-retest reliability of the SWLS. Over various lengths of time, scores appear fairly stable in test–retest correlations, ranging from .79 to .89 over a 2-weeks to 2-months period (Pavot & Diener, 1993), and also over longer time periods (Ehrhardt, Saris, & Veenhoven,

2000; Fujita & Diener, 2005; Schimmack & Oishi, 2005). This high reliability is perhaps to be expected from a measure that is meant to capture satisfaction over one's *life*.

Based on a series of experiments, correlation for the test-retest BDI ranged from .96 (n=46 student clients, see: Sprinkle et al., 2002) to .93 (n=26 outpatients over a 1-week interval, see: Beck et al., 1996).

The goal of the present study is to evaluate the relative measures of test-retest reliability in comparison with the frequently used affect measures described above, for which reliability assessments have been made in the past. In addition, the ERT will also allow us to investigate the relative persistence of negative and positive emotions.

4.2. Methods

a. Participants

The ERT scale was given twice to a group of 119 undergraduate students from the University of Warwick. Seven failed to complete the first or second test. The remaining 112 students completed both the test and retest and are included in the analysis below. Students were compensated with course credit. The test-retest scale was approved by the University of Warwick's ethics approval board. Participants reported as female in 90 cases (80.35%) and as male in 22 cases (19.64%). The mean age of participants was 19 years (M=19.08; SD=1.08) ranging from 18 to 26 years. Table 4.1 presents the language demographics showing 71.4% of the participant reported English as the first language.

b. Instruments

The Emotional Recall Task (ERT) 1.0. The ERT 1.0 contains two parts. In the first part, participants were instructed to write ten distinctive words describing their experienced emotions during the last month. In the second part of the test, each of the participant's responses are presented to the participant and the participant is instructed to rate "how often" they experienced each of their

emotions on a scale from 0 to 100, where 0 is “not often” and 100 is “very often”.

PANAS. The PANAS scale contains two 10-item emotional checklists, one for positive and one for negative emotion. Each item is rated for “the extent you have felt this way” in the past month on a Likert scale from *not at all* (score 1) to *extremely* (score 5). For example, people are asked to rate their responses to words like *jittery*, *interested*, and *strong*. The scales are shown to be highly internally consistent, largely uncorrelated, and stable at appropriate levels over a 2-month time period (Watson et al., 1988).

Diener’s Satisfaction with Life Scale (SWLS). The SWLS contains five agreement and disagreement statements on a 7-item Likert scale from *strongly disagree* (score 1) to *strongly agree* (score 7). For example, the SWLS contains questions such as “In most ways, my life is to my ideal” and “I am satisfied with my life.” This scale is reported to have high internal consistency (Cronbach alpha=.87, Diener et al., 1985).

UK Office for National Statistics Well-Being Measurement (ONS). The ONS *(ONS, Tabor & Stockley, 2018, May 17)* asks people to respond to four questions, each referring to a different dimension, regarding personal well-being on a scale from 0 to 100 where 0 is *not at all* and 100 is *completely*. For example, participants respond to “Overall, to what extend do you feel the things you do in your life are worthwhile” and “Overall, how happy did you feel yesterday”.

Beck Depression Inventory (BDI). The BDI (Beck et al., 1996) is a 21-item format questionnaire with four options for each item, with the scale ranging from not at all (0) to severe (3) for each item. For example, participants must choose between “I do not feel sad”, “I feel sad”, “I am sad all the time and I can’t snap out of it” and “I am so sad and unhappy that I can’t stand it”.

c. Procedure

Participants were invited to participate in an online study where they would be provided with a set of survey questions, on two separate occasions. They were asked to provide a matching identifier in both tests that could be used to match responses for each individual between the two separate occasions. Using the

Qualtrics platform, the ERT scale was provided together with other scales. Participants' first consented to being involved in each of the studies, separately for each occasion. After giving their consent, participants were expected to complete the ERT scales, followed by the other emotion scales and demographics information. The ERT scale was presented first so as not to prime individuals with emotion words. Following this, the rest of the scales were randomly presented. Participants received the link for the second survey 14-days after completing the first.

Table 4.1. Distribution of first language speakers among the participants

First Language	Number of participants	% of participants
English	80	71.4 %
Chinese	13	12.0 %
French	3	2.6 %
Punjabi	2	1.7 %
Tamil – Gujarati	2	1.7 %
Lithuanian	2	1.7 %
Chichewa	1	< 1 %
Finnish	1	< 1 %
Greek	1	< 1 %
Italian	1	< 1 %
Korean	1	< 1 %
Romanian	1	< 1 %
Serbian	1	< 1 %
Slovak	1	< 1 %
Spanish	1	< 1 %
Turkish	1	< 1 %

4.3. Result

The result revealed that 191 out of 1100 reported emotions do not have a valence value in the Warriner et al. (2013)'s norms. Some examples of unlisted emotion are *paralyzed, betrayed, stressed, lethargic, dedicated, down*, and *respected*. In addition, with 190 words that appear only once, the data contains 309 unique words. Because this study took place in parallel with the development of the ERT 2.0 (described in Chapter 3), it unfortunately does not have participants' unique word value ratings.

a. Test-retest reliability

Table 4.2 shows the test-retest reliability results for each of the tests. For purposes of comparison with prior work, we split the data into groups who completed the retest in the week following receiving it ("within three weeks") and those who took longer ("> 3 weeks"). Because only the ERT fulfilled the requirements of the Shapiro-Wilk test for normality, we use the non-parametric Spearman rank correlation for all the tests. For completeness, the Pearson product moment correlation for the ERT was .46 (.30-.59, 95% CI). Table 4.2 shows that the various scales all have comparable reliability. Moreover, the scores from scales in prior work are aligned with those found here (e.g., PA, .58, $p<.05$ and NA, .48, $p<.05$, from Watson et al., 1988; SWLS, .82 from Diener et al., 1985; BDI, .93 from Beck et al., 1996).

Note that the 95% confidence interval for each of the tests overlaps with the Spearman rank correlation for the ERT. The only exception to this is the PANAS negative correlation for the retest ("> 3 weeks"), which does not overlap with the ERT reliability rating. This is a curious because the PANAS negative correlation increases as the time since the original test increases. However, the second PANAS negative emotion test is still within the 95% confidence interval of the first. Nonetheless, the results confirm that the ERT is well positioned among the other scales in relation to test-retest reliability.

Table 4.2. Normality test and test-retest reliability between all comparable emotion scales

Emotion Scales	**Retest within 3 weeks**		**Retest > 3 weeks**	
	Correlation Index (Spearman Rank, *n*=88)	**95% CI**	**Correlation Index (Spearman Rank, *n*=24)**	**95% CI**
ERT	.48	.29, .62	.25	-.31, .64
PANAS Positive	.59	.41, .74	.51	.06, .85
PANAS Negative	.51	.32, .66	.64	.27, .87
Satisfaction with Life Scale	.68	.51, .81	.52	.13, .80
ONS Life Satisfaction	.44	.23, .61	.31	-.13, .66
ONS Life Worthiness	.44	.23, .63	.23	-.18, .59
ONS Happy	.33	.12, .51	-.04	-.49, .38
ONS Anxiety	.42	.21, .61	.13	-.34, .57
Beck Depression Inventory	.58	.41, .74	.41	-.05, .75

Because the PANAS and ERT scales request that participants evaluate their emotions over the past month, the reliability is partly a measure of

similarity between emotional experiences during the test-retest interval and the month prior to the initial test. This is supported by an analysis of the ERT correlations for individuals who completed the retest after different intervals, which goes down for the longer interval. For the ERT, "within 3-weeks" interval (n=88): rho=.48, 95% CI=(.29, .62) and for the "> 3-weeks" interval (n=22): rho=.25, 95% CI=(-.31, .64). Thus, a lower reliability score may also indicate a greater sensitivity to changes during the interval.

Table 4.3. Frequently reported emotion

Emotion	Frequency	Proportion
Happy	67	0.60
Excited	39	0.59
Stressed	36	0.50
Tired	26	0.39
Sad	24	0.38
Anxious	21	0.34
Confused	9	0.22
Content	8	0.21
Bored	7	0.19
Angry	6	0.19

b. Repeated emotions

The emotion words that participants were most likely to repeat are shown in table 4.3. To evaluate whether people were more likely to repeat emotions that they reported feeling more often in test 1, we determined which of the words participants repeated between test 1 and test 2. Figure 4.1 shows the distribution of 'how often' ratings for the repeated and non-repeated words. A comparison of the 'how often' distributions finds that participants were more likely to repeat words which they reported having felt more often, $t(708)=3.22$, $p<.01$. Similar results are found if the 'how often' ratings are first averaged within individuals (results of a paired t-test between each participant's repeated and non-repeated emotions, $t(104)=-4.038$, $p<.001$).

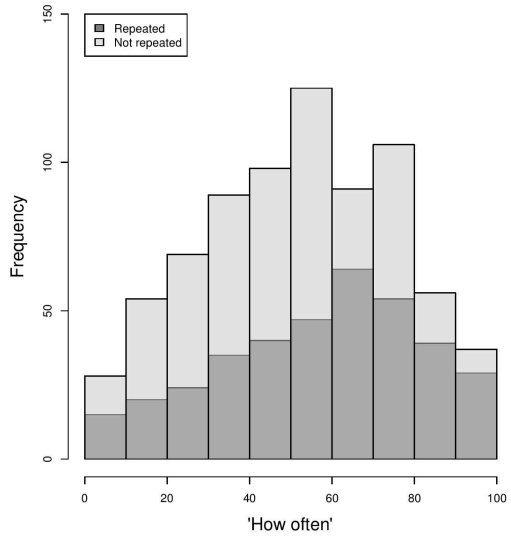


Figure 4.1. Distributions of repeated and non-repeated emotions

c. Were participants more likely to repeat positive or negative words?

Figure 4.2 shows the valence of the repeated and non-repeated words. The words that participants repeated were more positive than the words they did not repeat (results of a paired t-test between each participant's repeated and non-repeated emotions, $t(98)=-2.02$, $p=.046$. If results are not first averaged within participants, then there was no difference in valence between repeated and non-repeated words ($p=.43$).

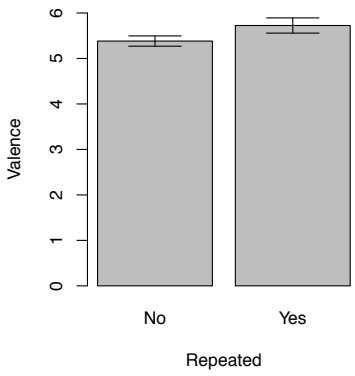


Figure 4.2. The mean valence (averaged first within participants) for repeated and non-repeated emotions

Note that the most frequently repeated emotion is *happy* (Table 4.3), which is also a frequently reported emotion in general. Indeed, *happy* is the most frequently reported emotion for both the test and retest, with 84 participants reporting *happy* in the test and 82 reporting *happy* in the retest. Thus, the bias towards positive emotions may be due to a social norm for reporting *happy* even when this emotion is not often felt. Figure 4.3 shows the distribution of 'how often' ratings and the relative position of the rating for *happy* (5.83). This supports the suggestion that *happy* is not reported most often because it most often felt, or because it is most persistent, but rather because it may be a social norm to report it.

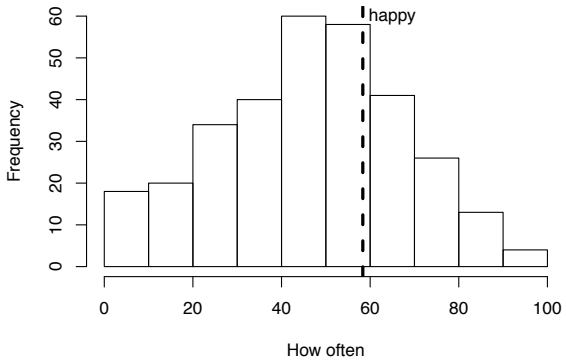


Figure 4.3. The relative distribution of 'how often' ratings for the initial test across all words. The rating for happy, the most frequently reported emotion and also the most likely to be repeated, is shown for comparison.

4.4. Discussion

Test-retest reliability for emotional scales is a standard measure for determining how well a test captures potentially persistent emotional states over a specified span of time. In this chapter, we investigated the test-retest reliability for the ERT over a period spanning from two weeks to one month. We found that the ERT yielded correlations on par with existing emotion scales. The highest reliability was when participants completed the retest within three weeks after the initial test (Table 4.2), with reliability scores going down after that time. All but one test were within the 95% confidence interval of their own retest and of

the other comparison scales. The results indicate that the ERT shares emotional test-retest reliability with other existing scales.

The repetition of emotions suggests that people are more likely to repeat emotions that they often feel. However, notably, *happy* is the most repeated emotion but is not an emotion that people were likely to report having felt frequently ('how often'). This may indicate that some emotions (like *happy*) are not reported because they are often felt, but rather because there is a social norm to report them. The ERT avoids this issue in some respects by asking participants to report on 'how often' they have felt each emotion and also report on ten recent emotions. The correlations with other scales (see also Chapter 2 and Chapter 3) suggest that this is sufficient to overcome the social norms for reporting certain emotions.

Chapter 5

Self-Reflections on Memory Retrieval in the Emotional Recall Task

5.1. Background

What comes to a person's mind when they are thinking about their recent emotions? This is relevant to the Emotional Recall Task because the way people search memory is known to influence what they find (Hills, Todd, & Jones, 2015). When searching for emotions, people may try to first recall past events that were memorable and then attempt to recall the emotions associated with those events. On the other hand, people may first try to recall an emotion that comes to mind, and then check to see if they have had that emotion. It is also possible that people are able to recall emotions directly, searching in some form of emotional memory space, that can be accessed through availability. This chapter aims to differentiate between these alternatives by using an individual interview format (the individual in-depth technique). Before addressing how this study was undertaken, the remaining portion of the introduction will address previous theory in this area.

New information is woven into old memories over time. Long-term memories are formed by a process of semantic encoding, whereby the meanings of new information create associations with pre-existing memories. The general theory of long-term memory retrieval demonstrates this semantic encoding using associative memory models (Raaijmakers & Shiffrin, 1981; Hills et al., 2015), whereby probes to memory are able to retrieve memories via activation based on associative relationships.

In the domain of emotions, a probe to memory like "think of emotions that you have experienced recently" relies on participants being able to use a probe like 'emotions' to recall specific emotions. To my knowledge, there is no previous discriminating among the various descriptions above in relation to how people recall emotions. Memory is probed by the activation of some mental stimulus, which then leads to an associated retrieval activation (Raaijmakers & Shiffrin, 1981). The probe in emotion recall may therefore be a vague but general 'emotion' probe, which activates a recent emotion (such as 'happy') without first recalling the context of that feeling. Alternatively, the probe may be event-driven, by first recalling an event that had an emotional signature, and then recalling the emotion associated with that event. We refer to these two kinds of search probes as emotion-driven and event-driven, respectively. As noted above, it is also possible that people have a strategy whereby they first recall an emotion from a list of emotions, and then ask if they have felt that emotion. We can call this list-driven emotion memory.

Past research has demonstrated that events that are associated with emotions are more likely to be recalled than non-emotional events. For example, in a study by Yuille and Cutshall (1986) contended that an emotional event experienced in real life encoded qualitatively different memories than a harmless laboratory event. Clifford and Scott (1978) also observed that people, even when in the laboratory, had better memory for emotional (violent) events than non-emotional (non-violent) events. In each of these studies, the results were interpreted as the result of emotions leading to the encoding of better memories.

Flashbulb memories are another example where memory may be enhanced due to emotions. Flashbulb memories are characterized as memories that are encoded during a particularly traumatic or surprising episode (Brown & Kulik, 1977). Though the standard belief was that flashbulb memories led to more accurate memories than non-flashbulb memories, this result has been questioned by more recent research. For example, a study by Talarico and Rubin (2003) studied the memories of individuals immediately after the September 11, 2001 terrorist attacks. The results of this study showed that memories for events associated with the terrorist attack (i.e., the flashbulb memory) were not more accurate than memories for non-flashbulb memories for more everyday activities. This led Talarico and Rubin (2003) and others (e.g., Hirst et al., 2015)

to suspect that flashbulb memories may be associated with other cognitive features (e.g., enhanced confidence) but they do not appear to be associated with more accuracy for event details. But, of course, the flashbulb status of memories is based on emotional relevance (Brown & Kulik, 1977), which implies a categorical change in even perception even if accuracy is unaffected.

Emotional mechanisms of memory may work by revisiting information for a certain time in association with internal states, either emotional or motivational, before becoming part of long-term memory storage (see: Papez, 1937; Hebb, 1949; Kornhubber, 1988). An improper emotional embedding of information may result in the inability to successfully store new information (Sapolsky, 1996), this phenomenon is called *mnestic block syndrome* (Markowitsch et al., 1999). For example, a 23-years old man lost his ability to store new and to retrieve old memories after he had seen a fire in his house. This is happened because he, as a child, had seen a man burning to death in a car, and this event seems to have induced a shock (mnestic block) condition resulting in observed amnestic condition. Personality characteristics may also affect emotional recall. People with repressive coping style, defined as low self-reports of anxiety together with high scores on defensiveness, have poorer recall of emotional experiences (Weinberger, Schwartz, & Davidson, 1979).

However, the key difference between the ERT and the studies described above is that the ERT is a recall-based task of emotions, not a recognition based task that asks one to recall emotions after being given the relevant event or an emotion from a list. In other words, flashbulb memories may be more memorable in relation to free recall—they may come to mind more easily than less emotional events. A potential example of this is the enhanced free recall of negatively valenced words relative to neutral words in a list-memory recall (e.g., Kensinger & Corkin, 2003b). Moreover, neutral words paired with high arousal taboo words also show an enhanced memory benefit compared to neutral word pairings (Guillet & Arndt, 2009). This may suggest that events associated with emotional experiences are more likely to be recalled.

A theory supporting this relationship is priority-binding theory (MacKay et al., 2004). Priority binding theory states that when arousing (i.e., emotional) events take place, they prioritize binding with more peripheral information that is occurring at the same time. Note this may also explain flashbulb memories. If

this is the case, then emotional experiences may drive better (broadly-defined) memory for the contexts in which they are taking place.

The observation that emotion-laden events are better recalled however does not address the question of whether the memory for that event is emotion-driven or event-driven. This is relevant to the emotional recall task, as the search process may indicate why people recall some emotions but not others. For example, some of our participants recalled emotions in the ERT that were substantially different from what one might infer from their PANAS score (see Figure 2.3 in Chapter 2). If an individual's PANAS indicates that they are experiencing highly negative emotions, but then their ERT leads to the production of a number of positive emotions, on what basis is the recognition versus recall of emotions producing different results? The PANAS is a recognition-based scale, which functionally uses a list-driven emotional memory format. This potentially misses emotions that participants are feeling most often (e.g., low arousal emotions). A better understanding of how emotions are recalled, for example, whether people use emotion or event-driven processes may be useful in guiding future research to understand how to help people better access their emotions.

In this chapter, we address this issue using an individual in-depth interview process. After recalling emotions from the ERT, participants are asked what came to mind when they recalled each emotion. Because this research is exploratory, following the interviews, the results are analysed using a qualitative framework to identify consistent patterns of recall across participants. However, this work also preserves the emotion-driven versus event-driven categorization. No participants reported using list-driven emotional recall.

5.2. Methods

a. Participants

The ERT interview was conducted on 15 graduate students at the University of Warwick. Participants were 15 interviewees randomly chosen to ensure the acceptable minimum participants in interview research (see: Guest, Bunce, &

Johnson, 2006; Crouch & McKenzie, 2006). Participants reported as female in 7 cases (46.67%) and as male in 8 cases (53.33%). The mean age of participants was 23 years (*M*=22.73; *SD*=2.46) ranging from 20 to 30 years. 4 of the participants were PhD students (26.67%), and 11 of the participants were master's students (73.33%).

b. Instruments explanation

The Emotional Recall Task (ERT) – Interview Version.

The Emotional Recall Task – interview version contains 4 sections.

Section 1 started with informed consent and developing rapport with the participants.

Section 2 contains the Emotional Recall Task. Participant was asked to write down 10 emotions describing their feelings in the past month. Participants were then asked to rate each of the emotions with the instruction: *Can you tell us how often you have experienced each of those emotions in the past month on a scale from 1 to 100?*

Section 3 contains four interview parts. In part 1, the interviewer probed each of the emotions by asking: *Before you said this* [emotion word]*, can you explain what came to your mind?* After probing all the words, the interviewer confirmed the participants responses by asking: Is [the response] *really what come to your mind before you said* [emotion] *or are you just thinking about it now because I am asking?* The interviewer then confirmed the time frame (part 3) by asking: *Are you sure this event occurred in the past month?* The interviewer then asked the participant to rate the emotions they produced again by asking: *If I ask you to* rate [emotion word] *again on a scale from 1 to 100, how frequently did you experience those emotions in the last past month?*

Section 4 contained the debriefing.

c. Interview Procedure

The ERT interview was conducted by using the individual in-depth technique. The in-depth interview is meant to be a personal and encounter in which open,

direct, verbal questions are used to elicit detailed narratives and stories (Briggs, 1986). The interview started by developing rapport and asking participants consent to participate in the interview. The interviewer explains the aim of the study, followed by what the participant needs to do during the interview, then explains that the study will take approximately 20 minutes to complete. Participants' informed consent was taken after explaining the study was completely voluntary, no specific personal information will be collected, and termination can be done at any time. Participants were then asked to complete the ERT by writing the emotion and rating each of the emotion on 1 to 100 scales. After completing the ERT, participant was asked the 4 parts of the ERT interview. After completing the ERT interview, the participant was debriefed and the interview process was complete.

5.3. Results

a. Comparisons of emotions before and after the interview

To evaluate whether the interview may have changed participants emotions, we collected frequency (how often) scores before and after the interview. If the interview influenced participants' emotions, then we might expect these ratings to be adjusted by the interview process, which might throw the interview into question. However, the average change in participants frequency ratings before and after the interview process was M=-1.4 and not significantly different from 0 (t(14)=-.90, p=.38). This suggests that participants' recollection of the intensity of their emotions were not changed by being directly questioned about those emotions. Figure 5.1 present the difference calculation between first and second score.

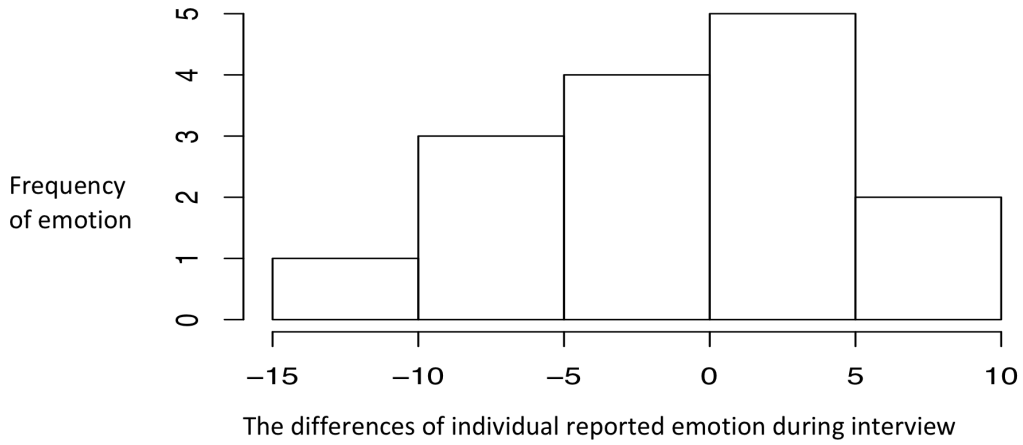


Figure 5.1. Histogram of differences (on a 100-point scale) in reported frequency of emotions before and after the interview

b. General observations

Based on what people said, there are two broad dimensions along which most of the emotions are recalled. One is time-oriented recall. People may recall emotions by focusing on past or future experiences. Eight participants from this study recalled emotions by focusing on past experience. For example, participant 2 came back from a trip and every emotion expressed was related to that trip. Participant 2 reported *"I feel tired because I came back from Scotland yesterday", "I am grateful because I finally had the courage to do a solo trip to Scotland", "I am passionate, I feel like I got the post vacation spirit", "I also feel sleepy and have a little bit of a migraine, because I did not have enough sleep during my Scotland trip".*

Two participants recalled emotions by focusing on future expectations. For example, participant 4 reports anticipating a trip to Italy, and expressed emotions related to that trip: *"I feel relaxed because I am going to Italy", "I wrote interested because I cannot wait to take lots of pictures there", "I am craving Italian pizza and pasta".*

Five participants recall emotions by focusing on the past and the future. For example, participant 13 mentioned both past and future oriented emotions: *"I feel active since I started going to the gym more often", "I feel relaxed after the*

Newcastle trip last weekend", "Zest! I am ready to start the Monday class", "I am hungry. I plan to go buy Subway in the Cannon Park".

Another dimension of emotional recall is the number of events that trigger the emotions. Four participants had all emotions triggered by a single event. For example, participant 6 was going to celebrate a birthday, and every emotion expressed refers to that birthday: *"I feel festive because this is my birthday month", "I am going to celebrate my birthday with my family", "I am happy because I will receive lots of birthday gifts"*.

Four participants had emotions triggered by two events. For example, participant 9 recalled emotions related to his work and his girlfriend such as *"I feel optimistic, I am going to apply for a new postdoc job", "I am also a bit busy because I start my teaching job this term", "I am in love, I am going to meet my girlfriend", "I feel lovable because of my girlfriend"*. Seven participants had emotions triggered by three or more events. For example, participant 1 reported all of the following: *"I am hopeless, my parents want me to take this course which I am not really keen on", "Well, I am also happy because they say education will bring good things soon after the exam, I can travel to Dubai", "I feel joyful, I met my friend for life here", "I am broke, I want money. Haha"*.

Participants also reported two different search processes used to trigger emotional memory. One is *event-driven*, in which participants recall the emotions based on a particular event, which then triggers the emotions associated with that event. For example, participant 1 explained that *"Yes, because my exam begun last month and since that time I am struggling a lot"*. Another example, participant 2 described that *"One month prior to my Scotland trip, I could not stop thinking about it"*. Both participants 1 and 2 chose a specific event then explored their emotions related to the event. Others searched over multiple events. For example, participant 7 reported the following: *"I am thinking about all the things I have done, and that is a lot, then I am thinking about what I feel towards all those things"*.

By contrast, some participants recalled emotional memories by retrieving the emotion they felt then thinking about what situation makes them feel that way. This is called *emotion-driven*. Participant 9 reported as follows: *"I am trying to ask my internal self about what is the feeling I have"*. Participant 10 reported the following: "*When you asked me, I asked myself what do I feel?*"

Participant 13 revealed his thought as follows: "*I just mention everything that came to my mind*".

These dimensions complement each other. In each case, participants report a temporal focus, the trigger, and an emotion or event focus. Table 5.1 presents the proportion of all dimensions.

Table 5.1. The proportion between interview variables

Search Probe	Time orientation	Event that triggered emotions	Number of participants	Examples of emotion	Example of probes
Emotion-driven	Past	single event	-	-	-
		two events	1	(participant 14) *I have something that makes me sad but I can't tell people; I procrastinate a lot because I cannot handle my problems.*	*those are feelings that is in my mind, I think. I just don't really think about it before you asked.*
		three or more events	3	(participant 10) *I hope no mistake or problem happened; I am grateful for what I have done in this job*	*when you asked me, I asked myself what am I feeling?*
	Future	single event	-	-	-
		two events	-	-	-
		three or more events	-	-	-
	Past + Future	single event	-	-	-
		two events	1	(participant 9) *I am also a bit busy because I start my teaching job this term; I am in love, I am going to meet my girlfriend*	*I am trying to ask my internal self about what is the feeling I have*
		three or more events	3	(participant 13) *I got small injured when in the football match last week; I* plan to go buy Subway in the Cannon Park	*I just mention everything came to my mind*
Event-driven	Past	single event	2	(participant 2) *I feel tired because I came back from Scotland yesterday; I am grateful because I finally had the courage to do a solo trip to Scotland*	*I keep thinking about when I was in Scotland.*
		two events	2	(participant 7) *My participants are not show up; I am proud of my partner's achievement*	*I am thinking about all the things I have done, and that is a lot, then I am thinking about what I feel towards all those things*
		three or more events	-	-	-
	Future	single event	2	(participant 6) *I am going to celebrate my birthday with my family; I am happy because I will receive lots of birthday gifts*	*I always thinking about my birthday*
		two events	-	-	-
		three or more events	-	-	-
	Past + Future	single event	-	-	-
		two events	-	-	-
		three or more events	1	(participant 5) *I missed my family; I want to sleep; I have the office on my own*	*I am giving those words by trying to remember everything happened around me*

Note: Event that triggered emotions refers to what all the emotions addressed. For example, if all emotions were focused on a single event, this is 'single event'. If all emotions focused on a future event, this is 'Future' time orientation. If some emotions are past and some are future, this is 'Past + Future'.

5.4. Conclusions

When people mention their recently experienced emotions, most try to retrieve information based on a recent situation they experienced. It may be related to something in the near future, something in the past, or something that happened near them. According to cognitive search theory, this may be a result of individuals using initial cues to memory that allow them to further elaborate on

specific events (Hills, 2006). Until they find the internal memory trigger, they keep searching their internal state for some hint of a past memory.

Once they find the internal memory trigger, many participants appeared to perseverate on that trigger—producing additional emotional memories related to a specific important event. Things that came to mind appeared to do so because of importance. When participants felt there were only one or two important things during the expected time frame, most of their emotions revolve around those things. If they cannot find any particular important thing during the expected time frame, they will retrieve emotions based on anything that come into memory. Important events appeared to have multiple emotions associated with them.

There are also differences in how people estimate the frequency of their recent emotions. People explained that probing about their emotions helped bring a deeper understanding of the reason they produced any particular emotion. Some of them then figured out that some of their emotions actually occurred more or less often than what they previously reported.

Chapter 6

ERT Cultural Project

6.1. Background

Cross-cultural similarities and differences within emotions, emotional expressions, and everyday conceptions of emotion are persistent questions in the study of emotion. How commonly are emotions felt across cultural groups and how invariant is the space of emotional expression? As each different language has different words for emotions, this raises a range of questions regarding how emotion is culturally varied. In comparing, for example, Americans' *polite* and Indonesians' *sopan*, we can ask to what extent these emotions are similar or overlap. As the vocabulary varies within the cultures, emotion words are likely to cluster differently in different cultures. Moreover, some emotions may be culturally unique, emphasized by some cultures but neglected by others (Levy, 1984; Lutz, 1988). For example, some research suggests that collectivist cultures tend to emphasize low arousal emotions, whereas individualist cultures emphasize high arousal emotions (Lim, 2016). Indeed, some cultures have no name for specific emotions recognized in other cultures (Lutz & White, 1986). For example, *haru* in Indonesian language refers to deep and prone feelings suddenly arising from either a neutral or pleasant stimulus in direct or indirect situation. In Google translate, the word *haru* will be translated as *emotion*. This translation is misleading because Indonesian has the word *emosi* for the word *emotion*, while *haru* is not a synonym of *emosi*. Another example is the word *kebelet*. *Kebelet* in the Indonesian language refers to a condition requiring immediate action or attention to either bladder control or marriage. In Google translate, *kebelet* will be translated as *have the need* (source: Badan Pengembangan dan Pembinaan Bahasa, 2018, October).

As it is disagreeable to provide emotional universals based on emotions from one language, ideally research should probe relevant issues in many different parts of the world so that a diverse database and set of interpretations of the data may be created. The present study, which examines the cross-cultural comparison of Indonesia and American emotional expression, is meant to contribute to that effort.

Home to about 300 ethnic groups and the world's largest Muslim community, there are a number of prior studies on Indonesian emotions. Bahasa Indonesia (the Indonesian language) has been decreed as the national language of the world's fourth most-populous nation since 1928 (Sneddon, 2003; Winskel & Widjaja, 2007), and is an interesting target domain for language-based emotion research (Anwar, 1980). Based on its linguistic typology, the Indonesian language is categorized as part of the Austronesian language (Winskel & Widjaja, 2007) used for centuries across a wide portion of Southeast Asia for business and trade. The emotional lexicon included within Bahasa Indonesia reflects the language's history. Most of the words are of Malay origin, and many are shared with modern Malaysians, but a few can be traced to Arabic (because of the influence of the Muslim religion), Dutch (because of hundreds of years of Dutch colonial rule), Japanese (because of Japanese control during World War II), and English (because of its worldwide cultural influence during the twentieth century) (Shaver, Murdaya, & Fraley, 2001). Bahasa Indonesia is also the official language of Indonesia, even though many Indonesians also speak a local dialect.

Interest on Indonesian culture and emotion was initiated by Ekman and Heider (1988). They conducted two experiments by asking Indonesian (Minangkabau) participants to judge emotions from photographs of Japanese-Americans and Indonesians (Minangkabau). Regardless of whether the person displaying the expression was male or female, American, Japanese, or Indonesian (Minangkabau), participants were capable in identifying *anger, disgust, happiness, sadness, fear,* or *surprise*. However, participants from different culture had difficulty identifying when expressions signalled other emotions such as scornful, haughty, smug, vain, or disdainful (experiment on Indonesian participants: Ekman & Heider, 1988; experiment on other culture participants: Ekman & Friesen, 1986). Another study used hierarchical cluster analysis to classify 124 Indonesian emotion terms. This examined how emotion

terms in Bahasa Indonesia corresponded to English emotional terms. For example, *love (cinta)* can be expressed in Bahasa Indonesia as *perasaan, getar hati, setia, edan, kasmaran, kangen, kemesraan, asmara, mesra, cinta, kasih,* and *sayang* (see: Shaver et al., 2001).

Studies on the reported experience of emotions produced between American and Indonesian, specific to the Minangkabau ethnic group, has been conducted to test the universality of emotions (see: Levenson, Ekman, Heider, & Friesen, 1992). The Indonesian experiments were conducted in Bukit Tinggi, West Sumatra, while the American experiments were conducted in Bloomington, Indiana. The American participants involved 62 subjects (27 men and 35 women, ages 18-30). Due to cultural constraints regarding contact between men and women, the Minangkabau experiment involved only male subjects (129 participants aged 16-27). All participants were asked to construct facial configurations for various emotional responses, like fear, anger, sadness, disgust, and happiness. The Minangkabau report of the target emotion occurred at significantly less than chance levels (z=-2.00, p=.02). In comparison, American is reported significantly greater than chance (z=5.76, p<.001). The Minangkabau were less likely to report experiencing the emotion that the facial configuration resembled than had Americans. For the Minangkabau, the task is missing a crucial element for emotional experience as defined by their culture, namely, the meaningful involvement of another person. As noted earlier, contrary with the Americans for whom the internal experience of emotion is very important, the Minangkabau focus primarily on the implications of emotion for interpersonal interactions and relationships (Heider, 2006).

The present study examines two main hypotheses based on the theory provided above; 1) Previous studies of the Minangkabau have found a strong proscription against negative emotion (Kato, 1982). If that sentiment is a general tendency of Indonesians and translates into the ERT, then Indonesians will produce more positive ERT scores; 2) With all the varied language that have been in contact over a long period of time and different ethnic linguistic groups**,** Indonesians will show greater variance in the emotions they recall.

6.2. The present study

We aim to extend the ERT research described in previous chapters to a cross-cultural comparison between American English and Indonesian speaking participants. Because ERT 2.0 is easily adapted to different languages and involves the recall of emotions in one's own language, this test can be described as 'content' neutral. Unlike the PANAS or other scales, which require the translation of emotional words and states initially chosen for a different culture, the ERT 2.0 allows participants to produce emotions specific to their own culture and, indeed, specific to each individual in the study.

6.3. Methods

a. Instruments Explanation

The Emotional Recall Task 2.0. The ERT 2.0 contains four sections. Section 1: participants are asked to provide 10 distinctive words describing their emotions in the past month. Section 2: the 10 words produced in Section 1 are presented in a random order with the instructions to rate how often have they experienced each of these emotions. Participants are then asked to give their ratings on valence (Section 3) and arousal (Section 4) by using a slider on a scale from 1 to 10. For the valence instruction, participants were asked to rate the pleasantness or unpleasantness of their emotion. For the arousal, participants were instructed to rate how calm or intense was the emotion.

The Positive and Negative Affect Schedule (PANAS). The PANAS (Watson et al., 1988) scale contains a 20-item emotional checklist, with 10 words provided for positive and negative emotion. Each word was rated on a Likert scale of *not at all* (score 1) to *extremely* (score 5). The scales are shown to be highly internally consistent, largely uncorrelated, and stable at appropriate levels over a 2-month time period (Watson et al., 1988).

Other Validation Scales. This study also involves well-being measurements to examine the generalizability of the ERT to other emotional aspects across cultures. Those scales include *Diener's Satisfaction with Life*

Scale (SWLS), *UK Office for National Statistics Well-Being Measurement* (ONS), and *Beck Depression Inventory* (BDI).

The SWLS (Diener et al., 1985) contains five agreement and disagreement statements on a 7-items Likert scale from *strongly disagree* (score 1) to *strongly agree* (score 7). This scale is reported to have high internal consistency (Cronbach alpha=.87). The SWLS items refer to how satisfied a person judges his own life in whatever criteria they choose, without specifying any aspects of life that may or may not contribute to the satisfaction level (Pavot & Diener, 2009). SLWS specificity is relevant for the cross-cultural research as specific items referring to life domains may be relevant to the satisfaction concept in some but not all cultures (Kuppens, Realo, & Diener, 2008). The ONS *(Tabor & Stockley, 2018, May 17)* asked people to respond to a four personal well-being questions on a scale of 0 to 100 where 0 is *not at all* and 100 is *completely*. The BDI (Beck et al., 1996) is a 21-item format with four options under each item, ranging from not present (0) to severe (3).

b. Experiment Procedure

The survey was set up in the Indonesia Language on the Qualtrics platform. We used the Indonesian version of PANAS, SWLS, and BDI that are commonly used in Indonesian research. The translation process of ERT and ONS involved a double translation process with two certified linguists. First, the English sentence was translated into the Indonesian Language. Then, without the English, the Indonesian translation would be translated back into English. The final translation of English was supposed to match the original English as closely as possible. The final Indonesian version of the survey has been approved by a certified Indonesian psychologist. After the study introduction and participants' informed consent, ERT was always presented first and randomly followed by the other scales and demographic information. The demographic questions included first language, country of origin, ethnicity, gender, age, and education.

c. Participants Demographics

The Indonesian version of ERT 2.0 was presented to Indonesian participants. After exclusion of participants who did not complete the survey and did not indicate Indonesian as their first language, the final sample consisted of 1259 participants (648 women (51.5%), 599 men (47.6%), and 12 refusing to report gender (< 1%). Thirty-nine (3.1%) of the participants were 18 years or younger; 938 (74.3%) were 19-30 years; 236 (18.7%) were 31-50 years; and 48 (3.8%) were 51 years or older. The mean age of participants was 28 years (*M*=27.6; *SD*=8.5) ranging from 13 to 69 years. The self-rated version of ERT were presented to 160 workers of Amazon Mechanical Turk (MTurk). Participants reported as female in 57 cases (35.4%) and as male in 103 cases (63.4%). The mean age of participants was 32 years (*M*=32.4; *SD*=9.7) ranging from 18 to 69 years. The MTurk data is the same as that presented in Chapter 3.

6.4. Result

a. ERT Scores

Participants' ERT self-rated is a weighted sum of the valence (or arousal) of emotion words combined with the frequency of each experienced emotion. The equation can be written

(1) $$ERT\ Valence\ Score = \frac{1}{n}\sum_{i}^{n=10}(V_i - 5.5)\ x\ W_i$$

(2) $$ERT\ Arousal\ Score = \frac{1}{n}\sum_{i}^{n=10}(A_i - 5.5)\ x\ W_i$$

Each of emotions with valence V_i (or arousal A_i) are weighted according to the reported frequency, W_i, from Section 2 of the Indonesian ERT. The frequency has a minimum of 0 and a maximum of 100, with higher scores meaning more frequently experienced. Table 6.1. presents the summary statistics of the ERT scores.

Table 6.1. Summary statistics of ERT scores

	Indonesian		American MTurk	
	Valence Score	Arousal Score	Valence Score	Arousal Score
Mean	44.22	54.29	91.36	27.24
Median	33.65	41.25	76.45	7.70
Standard deviation	158.39	125.22	162.00	133.70
Standard error of the mean	4.47	3.53	12.92	10.67
Minimum	-450.00	-302.35	-368.10	-301.05
Maximum	450.00	410.10	450.00	380.19

To provide the reader with a better sense of what the data looks like, Table 6.2 and 6.3 provide two examples from each culture, showing the participants with the most extreme ERT ratings and the words they produced. These extremes are valuable because they indicate that both cultures can plausibly experience high or low valence motions. Moreover, this is also a preliminary demonstration that the kinds of emotions that are seen as most positive or negative are shared in the two groups. We will revisit this claim when we discuss Table 6.7.

Table 6.2. Indonesian participants with lowest and highest ERT score

Lowest ERT score in Indonesian participants: -450

Indonesian Word	pahit	sakit	kecewa	sedih	pilu	gundah	muram	murka	murung	sepi
English Word	wretched	ill	disappointed	sad	miserable	upset	gloomy	wrath	moody	lonely
Valence	1	1	1	1	1	1	1	1	1	1
Arousal	9	9	4	4	2	5	5	6	2	10

Highest ERT score in Indonesian participants: 450

Indonesian Word	bersyukur	bahagia	senang	gembira	riang	cukup	ikhlas	berkah	tuhan	beruntung
English Word	grateful	happy	glad	joy	cheerful	content	sincere	blessed	god	fortunate
Valence	10	10	10	10	10	10	10	10	10	10
Arousal	9	8	8	7	8	7	8	8	8	8

Table 6.3. American participants with lowest and highest ERT score

Lowest ERT score in American participants: -368

Word	crappy	poor	lifeless	bored	tired	down	sad	angry	disgusted	disappointed
Valence	1	1	1	1	1	1	1	1	1	1
Arousal	5	5	5	5	4	5	6	9	6	5

Highest ERT score in American participants: 450

Word	happy	content	proud	excited	attentive	relaxed	calm	appreciated	accomplished	optimistic
Valence	10	10	10	10	10	10	10	10	10	10
Arousal	7	1	3	8	5	1	1	2	7	1

b. Cross-validation with other scales

Tables 6.4 and 6.5 show the correlations between the various scales. Table 6.4 provides the results for the Indonesian participants. Table 6.5 represents the results from the American MTurk sample previously produced in Chapter 3 but shown here for comparison purposes.

Both cultural groups show similar patterns of positive and negative correlations between the ERT and the other scales. There are no instances where a significant correlation in one culture has the inverse sign in the other culture. This suggests that the relationship between ERT and the other scales is preserved across Indonesian and American MTurk cultures.

The more subtle observation that more of the correlations are significant for the Indonesian versus the American MTurk sample is likely to be driven by sample size. The Indonesian sample is roughly an order of magnitude larger than the American MTurk sample (10 times the amount). Because a larger sample size increases the statistical power to detect significant relations where they exist, the correlations will tend towards greater statistical significance even if the effect sizes (the actual size of the correlation) are the same.

Table 6.4. Correlation matrix between all measures in Indonesian participants

	ERT_Valence	ERT_Arousal	Panas PA	Panas NA	SWLS	LifeSatisfy	LifeWorthy	Happy	Anxiety	BDI
ERT_Valence	1.00									
ERT_Arousal	0.34***	1.00								
Panas PA	0.45***	0.27***	1.00							
Panas NA	-0.15***	0.37***	0.08**	1.00						
SWLS	0.48***	0.19***	0.48***	-0.02	1.00					
LifeSatisfy_ONS	0.44***	0.25***	0.48***	-0.12***	0.63***	1.00				
LifeWorthy_ONS	0.48***	0.19***	0.50***	-0.17***	0.55***	0.69***	1.00			
Happy_ONS	0.50***	0.18***	0.48***	-0.20***	0.57***	0.66***	0.68***	1.00		
Anxiety_ONS	-0.13***	0.38***	0.04	0.62***	-0.03	0.01	-0.08***	-0.11***	1.00	
BDI	-0.16***	0.18***	-0.10***	0.49***	-0.02	-0.19***	-0.19***	-0.24***	0.37***	1.00

Notes: *p<.05,**p<.01,***p<.001

Table 6.5. Correlation matrix between all measures in mturk participants

	ERT_Valence	ERT_Arousal	Panas PA	Panas NA	SWLS	LifeSatisfy	LifeWorthy	Happy	Anxiety	BDI
ERT_Valence	1.00									
ERT_Arousal	-0.08	1.00								
Panas PA	0.53***	-0.03	1.00							
Panas NA	-0.25**	-0.02	0.14	1.00						
SWLS	0.55***	-0.03	0.63***	0.26***	1.00					
LifeSatisfy_ONS	0.56***	0.01	0.61***	-0.01	0.78***	1.00				
LifeWorthy_ONS	0.58***	0.03	0.64***	-0.12	0.63***	0.81***	1.00			
Happy_ONS	0.64***	0.01	0.59***	-0.08	0.63***	0.77***	0.78***	1.00		
Anxiety_ONS	-0.33***	-0.01	0.06	0.66***	0.13	0.04***	-0.04	-0.07	1.00	
BDI	-0.38***	-0.06	-0.12	0.71***	-0.02	-0.28***	-0.33***	-0.32***	0.54***	1.00

Notes: *p<.05,**p<.01,***p<.001

c. Frequently reported emotions

To further examine the cultural differences between the two samples, we examined the distribution of the most frequently reported emotions. The most frequently reported positive and negative emotions from the Indonesian and American MTurk samples are presented in Table 6.6 for comparison. The table shows the number of individuals who produced each word ('Frequency') and the mean valence for the word among those who produced it ('Valence') along with the 95% confidence interval for the valence ('CI'). Indonesian participants produced 1551 word types with 819 of those words only being produced by one participant. The American MTurk sample contains 623 word types and 415 words that appear only once.

The top 10 positive emotions from each cultural group show a high degree of overlap. Among the top 10, *happy, determined, calm, love* (or *loving*), are produced by both groups. The other words largely reflect similar semantic categories, with a few notable exceptions. The Indonesian sample produces *grateful, calm, sincere,* and *proud*, whereas the American MTurk sample produces *hopeful, energetic, and optimistic.* Though speculative, words like *grateful* and *proud* for the Indonesian sample and *hopeful* and *optimistic* for the American MTurk sample might reflect the past vs. future-oriented thinking that has been observed in prior work comparing non-Western and Western nations (Noguchi, Stewart, Oliva, Moat, & Preis, 2014).

The top 10 negative emotions share approximately eight words between the two samples: *sad, confused, angry, bored, anxious, worried, tired* (semantically similar to *exhausted*), and *scared* (semantically similar to *nervous*). The Indonesian sample produce the additional words *disappointed* and *annoyed*. The American MTurk sample, in comparison, produced *stressed* and *nervous.* Again, this difference may reflect a more general difference in past and future-oriented thinking, which may be a useful topic for future study.

Table 6.6. Frequently reported emotion

Indonesian top 10 positive emotion

Indonesian Word	English	Frequency	Valence	CI
senang	glad	667	7.30	0.21
bahagia	happy	577	7.35	0.23
semangat	determined	209	7.18	0.36
bersyukur	grateful	192	7.80	0.37
rindu	miss*	159	5.78	0.45
tenang	calm	123	7.16	0.49
cinta	love	117	7.50	0.49
gembira	joy	97	7.19	0.56
ikhlas	sincere	77	6.97	0.60
bangga	proud	70	6.80	0.62

American MTurk top 10 positive emotion

Word	Frequency	Valence	CI
happy	98	8.35	0.37
excited	30	8.40	0.48
calm	27	8.00	0.65
hopeful	25	8.16	0.44
good	20	7.75	0.65
determined	17	7.77	1.10
loving	17	9.06	0.51
content	15	8.93	0.65
energetic	13	8.00	0.89
optimistic	14	8.86	0.84

Indonesian top 10 negative emotion

Indonesian Word	English	Frequency	Valence	CI
sedih	sad	631	4.14	0.24
kecewa	disappointed	319	4.41	0.34
bingung	confused	308	4.67	0.33
marah	angry	303	4.34	0.35
lelah	exhausted	289	4.60	0.35
takut	scared	252	4.45	0.37
kesal	annoyed	203	4.51	0.45
bosan	bored	195	4.79	0.44
cemas	anxious	140	4.05	0.51
khawatir	worried	141	4.58	0.51

American MTurk top 10 negative emotion

Word	Frequency	Valence	CI
sad	48	4.17	0.86
tired	35	3.46	0.76
anxious	27	2.15	0.56
worried	20	1.85	0.61
angry	21	4.38	1.27
stressed	11	1.36	0.40
depressed	11	1.81	0.95
confused	11	2.73	1.03
bored	10	3.10	0.74
nervous	10	2.40	0.84

Note:* *Rindu* does not have a direct translation to its positive equivalent in English, but *miss* is the closest word in meaning.

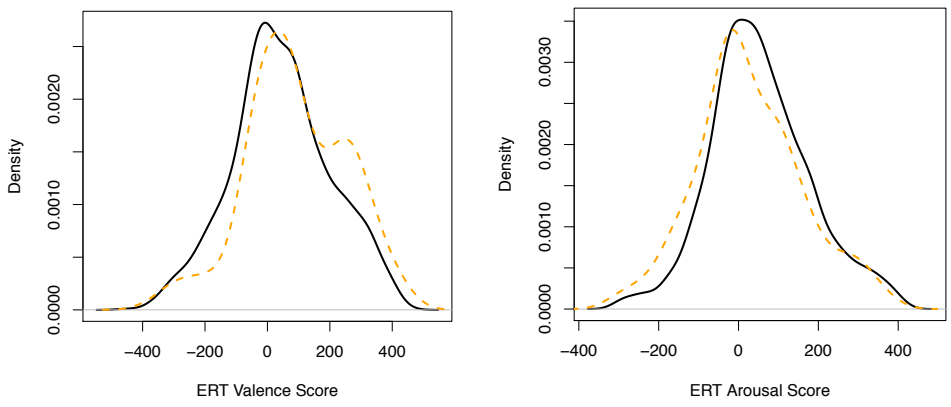


Figure 6.1. Density plot of ERT scores comparing Indonesian (solid-black line) and American MTurk (dotted-orange line) samples. The density plot shows the proportional distribution of ERT scores, with the area under the curve equal to 1. Each individual's ERT is represented by a small Gaussian kernel which is added for all participants and divided by the number of participants.

d. Difference in Emotion Distributions Between Indonesian and American Samples

Figure 6.1 shows the distributions of ERT scores for the Indonesian and MTurk samples. Because these samples are not normally distributed, we use a Kolmogorov-Smirnov test to determine the statistical likelihood that they are drawn from the same continuous distribution. The results of the Kolmogorov-Smirnov test indicate that this is statistically unlikely for valence (D=.14, p=.01). Visual inspection of the figure suggests that the American MTurk sample has a higher ERT score, and this is confirmed by inspection of the means as well (American MTurk: M=91.36 (SD=162.00), Indonesian: M=44.22 (SD=158.38); t(195.13)=-3.44, p<.05). The results of the Kolmogorov-Smirnov test indicate that this is also statistically unlikely for arousal (D=.13, p=.01, American MTurk: M=27.24 (SD=133.70), Indonesian: M=54.29 (SD=125.22); t(191.77)=2.40, p<.05).

Further inspection of the individual words (tokens) produced by the Indonesian and American MTurk samples show that the American sample produces more highly positive words and fewer highly negative words (see Figure 6.2). The Kolmogorov-Smirnov comparison on valence shows that these two distributions are also statistically unlikely to be drawn from the same distribution (D=.10, p<.001). As the figure confirms, the American MTurk sample has a higher mean as well (American MTurk: M=6.30 (SD=3.07), Indonesian: M=5.69 (SD=3.18); t(1838.8)=-7.00, p<.05).

The Kolmogov-Smirnov comparison on arousal (see also Figure 6.2) shows that these two distributions are also statistically unlikely to be drawn from the same distribution (D=.05, p=.006). As noted in the figure, the Indonesian sample has a higher mean on arousal (American MTurk: M=5.71 (SD=2.67), Indonesian: M=5.97 (SD=2.58); t(1784.6)=3.38, p<.05).

Finally, we can inspect the distribution of word types, which reflects the background emotional categories that participants appear to be selecting from, based on the samples they produce. This is presented in Figure 6.3 and further shows the mean valence for each word type from the Indonesian sample alongside the mean valence for each word type from the American MTurk

sample. Mean valence is computed by taking the mean of the self-reported valence for the word from each person who produced that word. Once again, the Kolmogorov-Smirnov test indicates that the Indonesian and American MTurk samples are unlikely to be drawn from the same continuous distribution (D=.12, p<.001, American MTurk: M=6.28 (SD=2.85), Indonesian: M=6.05 (SD=2.58); t(1096.9)=-1.78, p>.05).

Figure 6.3 also presents the mean arousal for each word type from the Indonesian sample alongside the mean arousal for each word type from the American MTurk sample. Mean arousal is computed by taking the mean of the self-reported arousal for the word from each person who produced that word. On arousal, the Kolmogorov-Smirnov test indicates that the Indonesian and American MTurk samples are unlikely to be drawn from the same continuous distribution (D=.07, p=.03, American MTurk: M=5.94 (SD=2.37), Indonesian: M=6.07 (SD=2.20); t(1120.9)=1.21, p>.05).

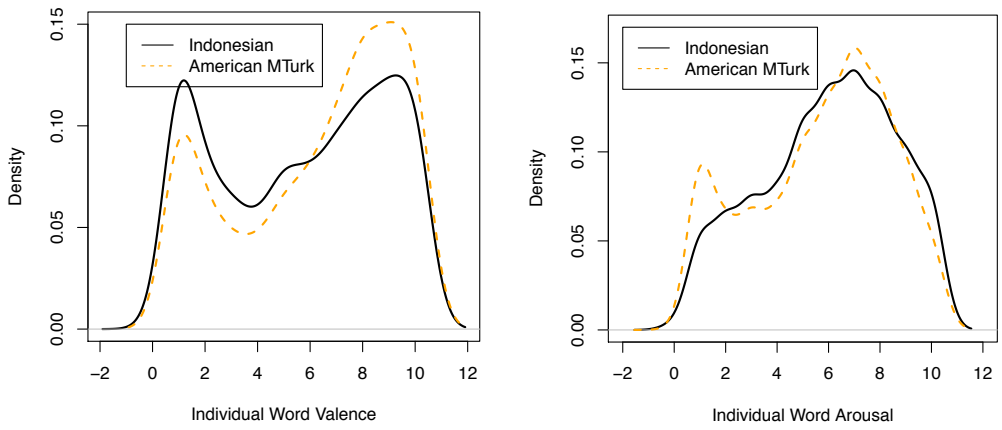


Figure 6.2. Distribution of word valence and arousal, comparing all words (tokens) produced from Indonesian (solid-black line) and American MTurk (dotted-orange line) participants. The density plot shows the proportional distribution of emotion words, with each word represented by a Gaussian kernel, with the area under the curve equal to 1, on valence (left) and arousal (right).

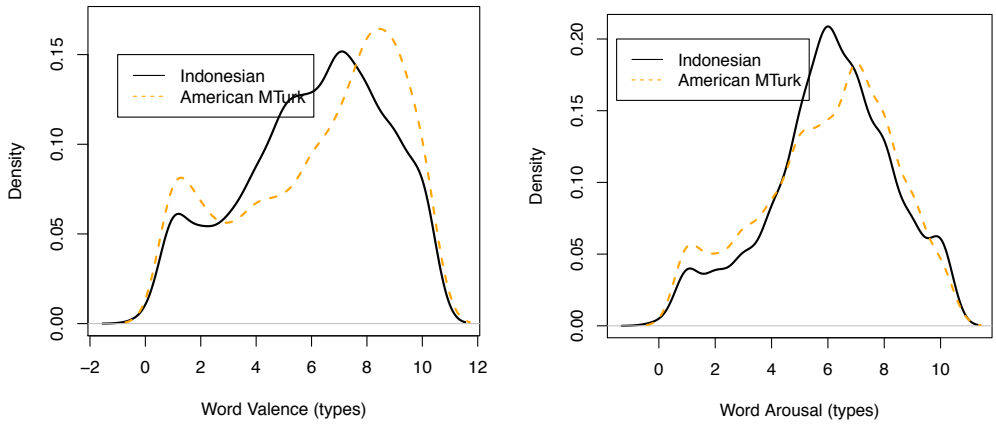


Figure 6.3. Density plot of word valence and emotion, comparing unique words (types) produced from Indonesian (solid-black line) and American MTurk (dotted-orange line) participants. These plots are computed as in Figure 6.2.

We further examine these cultural differences by examining the additional emotional scales. Figure 6.4 shows the distributions of scores for the Indonesian and MTurk samples on all comparable scales. The results of the Kolmogorov-Smirnov test on SWLS indicate that the difference in distributions is statistically similar (D=.11, p=.06). Visual inspection of the figure suggests that the American MTurk sample has a higher SWLS score (American MTurk: M=23.18 (SD=7.59), Indonesian: M=21.94 (SD=6.98); t(194.76)=-1.96, p=.05).

The results of the Kolmogorov-Smirnov test on BDI indicate that this difference is statistically unlikely (D=.16, p=.002). Visual inspection of the figure suggests that the Indonesian sample has a higher BDI score (Indonesian: M=36.08 (SD=10.98), American MTurk: M=35.43 (SD=14.25); t(183.75)=.56, p>.05). The Kolmogorov-Smirnov test on the ONS scale indicates that the distributions are statistically similar (D=.07, p=.40), with figure also suggests the Indonesian sample has a higher ONS score (Indonesian: M=259.03 (SD=63.00), American MTurk: M=248.119 (SD=71.69); t(191.51)=1.84, p>.05).

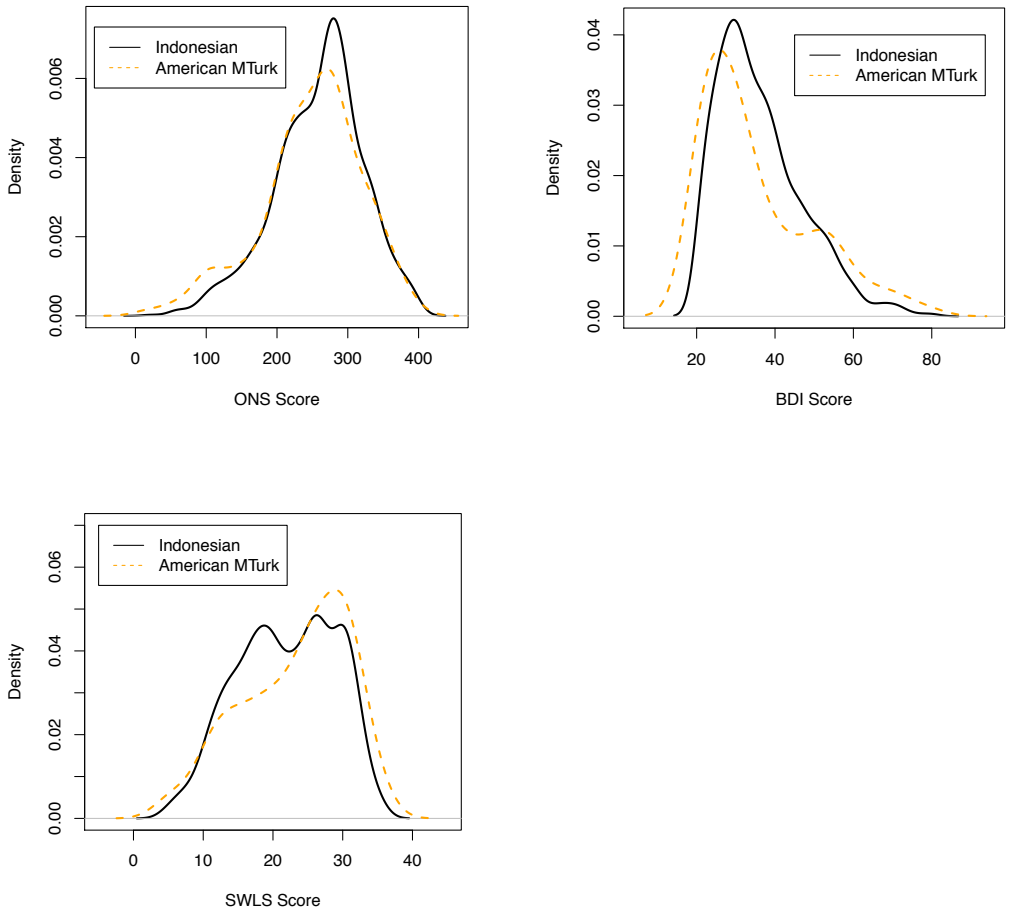


Figure 6.4. Density plot of all comparable sales comparing Indonesian (solid-black line) and American MTurk (dotted-orange line) samples.

The density plot shows the proportional distribution of each score, with the area under the curve equal to 1. Each individual's score is represented by a small Gaussian kernel that is added for all participants. ONS (top left), BDI (top right), SWLS (bottom left).

6.5. Discussions

This study investigated the applicability of the ERT to an Indonesian sample. This also allowed us to further compare the emotional expression of an Indonesian sample of that with an American MTurk sample. The principle result is that the ERT applied to an Indonesian sample shows similar correlations with other emotional scales that we see in a comparison of scales with the American

MTurk sample. Since the ERT self-rated version allows participants in any language to produce whatever words they like and is therefore content neutral with respect to emotional expectations, the ERT is potentially a valuable tool for studying emotional states across cultures. Our results provide evidence supporting this conclusion and therefore provide a content neutral tool for the easy collection of emotional states across cultures.

Do Indonesians produce more positive emotional words? Our results suggest the opposite. Even though both Indonesian and American samples can produce words at the extremes of the emotional scale, American participants, however, tend to produce more highly positive valence words. This was not expected based on inferences from Minangkabau culture (Kato, 1982). However, one potential explanation is that globalization has encouraged the younger generation to be more expressive, brave, and straightforward in expressing their emotions or thought (see: Heppell, 2004 and Suraya, 2003). This is a topic that would be ideal for future study in Indonesia, as it could be examined by looking at age-related differences in emotional expression in Indonesian culture and also examining between ethnic groups with more or less exposure to globalization.

Are Indonesians less likely to experience extreme emotions or arousal? The results suggest that this is the case. Note that in all cases, except for the valence of word tokens and ONS, the American MTurk sample shows a higher variance than the Indonesian sample. This indicates that the American MTurk sample is producing more varied emotions and arousal ratings. This is also evident in the density plots, where (excepting ERT and ONS) the American MTurk participants tend to produce more data at the extremes of the distribution. The observation that the tokens do not show these extremes for valence is suggestive that it is not the words that Indonesians can draw from that is the reason for this variance, but rather the words that most individuals choose to report their own emotions. The emotional choice set is sufficiently large to include extreme valence words, but Indonesians appear to be less likely to choose them.

An explanation for the potential lower ERT ratings of the Indonesian sample is also provided by the World Happiness Report 2018 (Helliwell, Layard, & Sachs, 2018). In the previous findings of ERT (for reviews see Chapter 2: ERT Validation), ERT showed a strong correlation with happiness in all the

comparison scales. According to the World Happiness Report 2018, Finland (happiness index=7.632) is the happiest country and Burundi (happiness index=2.905) is the least happy country. The Republic of Indonesia (happiness index=5.093) ranked 96th out of 156 countries while The United States of America (happiness index=6.886) ranked18th in happiness across countries. These overall ranking are based on the pooled result from the Gallup World Poll Surveys from 2015 to 2017 by asking people to evaluate the quality of their current lives on a scale of 0 to 10. The ERT ratings may therefore be picking up on a larger tendency for American's to have generally higher valence emotional states than Indonesians. Additional research based on this explanation could look at ERT scores across a broader sample of countries and people within countries.

Finally, we would like to highlight the preliminary observation that negative emotion in Indonesia may be related to past-oriented reflections on emotional state. The American MTurk sampled appeared to show more future-oriented reflections. This orientation of time-perspective has been observed to be related to economic outcomes (Noguchi et al., 2014) as well as emotional states (Wyrick & Wyrick, 1977). For example, Wyrick and Wyrick (1977) observed that depressed individuals tended to be more past-oriented than a similar control population. Our results suggest this may be observed at a national level, which is an exciting area for future research.

Chapter 7

The Emotional Literacy and Peak-End Rule: Supporting the ERT

7.1. Background

Previous chapters have established the ERT as a useful tool to measure emotional states. The present chapter extends this idea to evaluate a principle of experienced utility called the peak-end rule. The idea that utility is a property of experience that can be measured was initially proposed by Francis Edgeworth (1871). He proposed the idea of a *hedonometer*, which was a speculative instrument that could measure the ups and downs of one's experience in much the same way that a thermometer can measure temperature. If one imagines above which experiences are positive and below which experiences are negative, then if one takes the area under the curve of this measurement one can compute a measure of utility over time. In theory, a good decision maker should not be willing to pay two different amounts for the same experienced utility. The peak-end rule is the observation that people will, in fact, pay different amounts for the same experienced utility, because when they reflect on their utility they do not take the area under the curve, but rather evaluate their experience in relation to the *peak* experience and the *end* experience.

Redelmeier, Katz, and Kahneman (2003) demonstrated this is in experiment of memories about pain following a medical procedure. The procedure lasted from 4 minutes to 69 minutes with pain intensity varying from 0 ("no pain at all") to 10 ("intolerable pain"). Pain intensity was measured during

and after the procedure. We will compare two patient experiences, shown in Figure 7.1. Patient B's procedure lasted much longer than patient A. The last rating on the pain scale before the end of the procedure was 7 for patient A and 1 for patient B. Patient A retained a much worse memory of the episode than patient B, despite experiencing less overall pain (as measured by the area under the curve). Indeed, Redelmeier and Kahneman (2003) found that people's memory of the pain was best explained by their peak and end experiences, giving rise the *peak-end rule.* As they also explained, this meant that people tend to have *duration neglect* when evaluating experienced utility, such that the overall time is not taken into account.

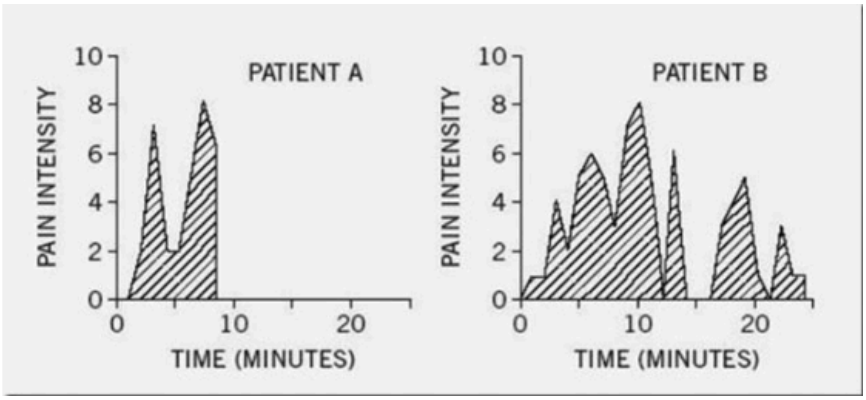


Figure 7.1. The differences between participants' assessment of pain experienced on varied time procedures

The procedure for patient A (left) was ended at the seventh level of pain, while the procedure for patient B (right) was ended at the first level of pain. This experiment proposed that one's experienced utility would vary as a function of time. (Taken from Kahneman, 2011, p. 379).

Both peak and end emotions are distinct in one's memory and are not attenuated by other emotions. This concept has been described as a *focusing illusion* (Schkade & Kahneman, 1998). There will be a powerful tendency to exaggerate the importance of any aspect of one's life when one focuses attention on it. Schkade and Kahneman (1998) asked groups of students from the Midwest and California where they would be happier if they lived, in California or the Midwest. Both student groups believed that living in California would make

them significantly happier. However, self-reported happiness was the same for students in both locations. Schkade and Kahneman (1998) explained that their participants tend to focus on weather when the contrast between the two choices highlighted weather. But when reporting on their own well-being, weather was not taken into account, hence the focusing bias.

The focusing bias may explain the peak effect in the peak-end rule, because when asked about the experience over a given period of time, the peak experience is likely to be the experience of highest contrast with other 'normal' experiences. This may influence the ERT, as some experiences are likely to be of higher intensity than others and may therefore have a stronger influence over the memory of one's emotions over a specific period of time. In addition, the 'end' experience is the most recent experience and therefore the most accessible to memory (e.g., Murdock, 1963). One may therefore hypothesize that the ERT—if it follows a similar cognitive process to evaluations of experienced utility—will show prominent effects of the peak and end emotions.

A second focus of this chapter is on emotional literacy. Chapter 6 showed that even though Indonesian's, as a group, had access to words that could express very high and very low valence words, they tended to produce them less often than Americans. This line of reasoning naturally leads one to ask if the store of one's own emotional lexicon influences the emotions that one recalls. By having people list all the emotions they can think of, we may therefore assess the extent to which their emotional lexicon can predict their ERT.

To measure each individual's emotional lexicon we introduce emotional literacy task. This is inspired by fluency studies like the verbal fluency test (see: Mueller et al., 2015; Wysokiński, et al., 2010). Verbal fluency evaluates one's ability to retrieve words in relation to a specific category (Lezak, Howieson, Loring, Hannay, & Fischer, 2004). Verbal fluency tests (also called category fluency tasks) measure fluency by having participants produce as many words as possible within a category over a fixed period of time. For example, listing as many animals as possible in one minute (i.e., the animal fluency task). The importance of this ability has been illustrated across various aspects of life including communication, social interactions, memory, occupational functioning, lower and higher order cognitive skills (Stolwyk, Bannirchelvam, Kraan & Simpson, 2015). Verbal fluency tests are also valid measures of vocabulary size,

inhibition ability and cognitive impairments (Shao, Janse, Visser, & Meyer, 2014). More recently, verbal fluency tests become instrumental in the development of cognitive models of memory search and the network structure of the mental lexicon (e.g. Hills, Jones & Todd, 2012; Zemla & Austerweil, 2018; Troyer, Moscovitch, & Winocur, 1997).

7.2. The present study

This chapter asks two questions. First, does emotional fluency predict emotional recall? And second, to what extent do peak and end emotional experiences explain the outcome of the ERT? To address these questions, this chapter introduces the emotional literacy task and a modification of the ERT to capture peak and end emotions. The emotional literacy task asks participants to produce all the emotions they can think of in a two minute period. The ERT is modified to include a measure of 'intensity' and also to allow participants to report which emotion was felt most recently. These are then used to evaluate the influence of the emotional lexicon and the peak-end rule on the ERT.

7.3. Methods

a. Instruments Explanation

The Emotional Recall Task (ERT). The ERT contains four parts. In the first part, participants were instructed to write ten distinctive words describing their recently experienced emotions during the last 30 days period. In the second part, all the emotion words produced by the participant were presented to them again and the participants were instructed to rate how often they had experienced each of those emotions using a 0 to100 slider. In the third part, all the words written were presented again and participants were instructed to rate how intense they had felt emotions on a scale of 0 (not at all) – 100 (extremely intense). In the last part, participants were asked to answer which emotion they had experienced most recently.

Emotional Literacy. In this part, participants were instructed to list as many distinctive words describing emotions as they could in two minutes. Participants type these words into a response window on the computer screen.

The Wong – Law Emotional Intelligence Scale (WLEI). The WLEIS is a 16-item self-report instrument for measuring emotional intelligence (Wong & Law, 2002), designed to match the four dimensions of emotional intelligence proposed by Salovey and Mayer (1990) – (1) self-emotion appraisal (SEA); (2) other's emotion appraisal (OEA); (3) use of emotion (UOE); and (4) regulation of emotion (ROE). Participants were instructed to indicate how often they experienced four corresponding statements on each dimension on a 5-point Likert scale with responses varying from "None of the time" (1) to "All of the time" (5).

UK Office for National Statistics Well-Being Measurement (ONS). The ONS was developed to assess personal well-being which capture three types of well-being; evaluative, eudemonic and experience. Participants were instructed to evaluate their overall life satisfaction, worthiness of things they do, happiness and anxiety on a 0 to 100 scale, with a total score calculated for each participant (Tabor & Stockley, 2018, May 17).

The Satisfaction with Life Scale (SWLS). The SWLS consists of five statements measuring the satisfaction with life. Participants were instructed to choose their agreement on a 7-point Likert scale from 'Strongly Disagree' (1) to 'Strongly Agree' (7). The total score is calculated for each participant with a higher score reflecting more subjective well-being (Diener et al., 1985).

b. Procedure

Participants were invited to take part in a short survey about emotions and well-being. Those who accepted the invitation were redirected to the survey on Qualtrics. On redirection participants were first shown an information page, provided with a detailed description on the length of the survey, its voluntary nature, assurance of confidentiality, and researchers' contact. Participants had the options to give consent or withdraw from the study. Following participants consent, they were then directed to the modified ERT, emotional literacy task,

WLEIS, ONS, and SWLS questionnaires respectively. In the modified ERT, the participants were asked to report the most recent emotion they experienced and the intensity of each emotion. Upon completion, participants were shown a debrief page restating the aims of the study.

c. Participants

2,000 participants were recruited via Amazon Mechanical Turk (MTurk) in return for a payment. They were required to be English speaking and based in the United States. All participants were paid £0.75 to complete the questionnaire which took approximately 6-7 minutes.

After data cleaning procedure, 5 responses were disregarded due to duplicate IP addresses; 132 responses were disregarded due to incomplete answers (either report less than the required number of ten recalled emotions, or left the questionnaire unfinished); and 228 responses were disregarded due to invalid answers (report numbers in the emotional fluency task). Subsequently, a final sample size of 1,635 participants was used for analysis. No demographic information was collected in this study.

7.4. Result

a. Descriptive Scores

Table 7.1 presents the descriptive statistics for the ERT scores, numbers of word reported in emotional literacy, WLEI scores, ONS scores, SWLS scores, and participants' peak-end emotions. The ERT scores were calculated as a function of the mean valence of the emotions recalled (for discussion see: Chapter 2 – ERT Validation). Participants' emotional literacy score were computed from the total number of emotions reported by each participant. Participants' WLEI, ONS, and SWLS scores were calculated by summing up the values of their answers. Valence for peak and end words was estimated for each participant using the Warriner et al. (2013) norms.

Table 7.1. Descriptive statistics between all measures

Measure	Mean	Median	Standard Deviation	Minimum	Maximum
ERT	45.79	51.17	90.74	-229.46	257.40
Emotional Literacy	14.14	13.00	6.43	0.00	60.00
WLEI	60.62	61.00	9.23	16.00	80.00
ONS	235.85	246.00	64.20	3.00	400.00
SWLS	22.93	25.00	7.64	5.00	35.00
Peak Valence	5.75	6.62	2.28	1.71	8.48
End Valence	5.88	6.70	2.35	1.79	8.48

b. Emotional literacy and the ERT

Did emotional literacy predict the ERT? As shown in Table 7.2, the number of words a participant produced in the emotional literacy task was not predictive of their ERT. This indicates that the size of the emotional lexicon is a poor predictor of general affect. However, this does not evaluate whether or not the structure of the lexicon itself is predictive of the ERT. Future work should evaluate the network structure of the emotional literacy task (following Zemla & Austerweil, 2018) and also the overall valence of the emotional lexicon to see if these are predictive of the ERT. This would not imply causality but may nonetheless inspire future experimental studies that can examine the influence of training in the emotional lexicon. Such a study would be further supported by the relationship between emotional literacy and the WLEI scale, ONS, and SWLS. These each show a small positive effect with emotional literacy. This indicates a relationship between the emotional lexicon and emotional intelligence and positive emotion.

c. Emotional Intelligence and the ERT

The relationship between emotional intelligence and positive emotion is also supported by the correlations between WLEI and the other emotional scales, including the ERT. In all cases, higher emotional intelligence was associated with greater positive emotion. Though, again, emotional intelligence was not well-predicted by emotional literacy, which suggest boundary conditions on what exactly emotional intelligence is composed of. WLEI questions are largely

focused on interpersonal and subjective emotional experiences, not necessarily on emotional specificity (as defined in Chapter 2).

d. Peak-end and ERT

Peak and end valence are both well-correlated with the ERT and also with the other emotional scales (see also Figure 7.2). This is already a promising result as the size of the correlations for the peak and end experiences are fairly large and suggest that a substantial amount of the ERT's variance may be explained by these two factors alone. We examine this further below.

Table 7.2. Correlation between all measures and participants' peak-end emotion

	ERT	Emotional Literacy	WLEI	ONS	SWLS	Peak Valence	End Valence
ERT	1.00						
Emotional Literacy	-0.01	1.00					
WLEI	0.42***	0.06*	1.00				
ONS	0.53***	0.07**	0.46***	1.00			
SWLS	0.57***	0.05*	0.43***	0.70***	1.00		
Peak Valence	0.62***	-0.03	0.28***	0.37***	0.41***	1.00	
End Valence	0.63***	-0.03	0.28***	0.40***	0.45***	0.56***	1.00

Note: *p<.05,**p<.01,***p<.001

Participants' peak and end emotions were highly correlated, results of a Pearson product moment correlation, *ß=.56 (.52-.59, 95% CI)*. This includes participants for whom peak and end emotions are the same. Excluding the analysis to participants with different peak-end words reduces this effect to *ß=.32 (.26-.37, 95% CI)*.

Do peak and end emotions independently explain the ERT? A multiple regression was carried out to investigate the predictive power of the peak and end emotions. A significant regression of equation was found, $F(2,1632)=814.3$, $p<.001$, $R^2=.50$. Both peak and end were independently predictive: the peak emotion: $ß=15.74$, $p<.001$, and end emotion: $ß=15.68$, $p<.001$. Once again, multiple regressions were carried out after excluding all the cases with identical peak and end words. A significant regression equation was found ($F(2,1097)=680.3$, $p<.001$), with an R^2 of .55. For the peak emotion: $ß=18.58$, $p<.001$, and for the end emotion: $ß=17.99$, $p<.001$. These results indicate that

both peak and end emotions contribute independently to our recent emotional memories over the past month.

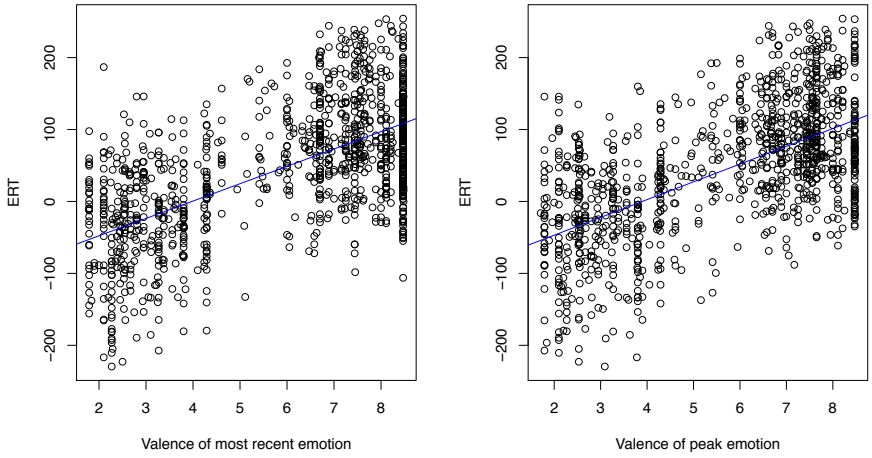


Figure 7.2. Distribution of participants' responses between ERT scores and peak emotion (the highest intensity) (right); between ERT scores and the most recent emotion (the lowest intensity) (left) and ERT score. Blue line is calculated from the regression between peak and end emotions

7.5. Conclusions

How people recall emotions is potentially constrained by their mental lexicon. In addition, what we recall is likely to be influenced by our most recent experiences and the peak experiences. The results presented here do no entirely support the first claim, but they do support the latter claim.

The results find no relationship between the putative size of one's emotional lexicon, as measured by the emotional literacy task, and one's ERT. This may be due to a limitation of the emotional literacy task or because the important information in the emotional literacy task is contained in its relationship with the structure of the mental lexicon. Future research will be needed to address this.

The peak emotion and the most recent emotion both independently predicted the ERT scores for individuals. The coefficients reveal that they have approximately the same effect, with an increase in peak or end emotion reflecting an increase in the ERT. Together the R^2 indicates that they explain about half of

the variance of the ERT, which is substantial since they only represent two of the words from the ERT.

Chapter 8

Conclusion

How our emotional state is determined is a consequence of both how we experience our lives but also, as demonstrated by the peak-end rule, how we recall those experiences. This thesis was dedicated to investigating this recall process to help us better understand our emotional states. This chapter provides brief summary of the central contributions of this thesis.

Chapter 1 explained the motivation for developing the Emotional Recall Task as a new recall-based emotion measure. This chapter described several established emotional scales with their pros and cons in emotion measure and explained how these scales might miss characteristics of emotional experience such as emotional specificity and emotional breadth. Notably, it argued that differences in specificity and breadth might explain the discord in previous attempts to define a specific set of universal emotions.

Chapter 2 explained that existing self-report based emotion scales typically involve recognition of emotions from a predetermined emotion checklist. This led to the proposal for an emotion scale that relies on recalled memory. Instead of asking people to evaluate their emotional experience in relation to a list of terms that may nor may not properly cover their emotional experience, we asked people to produce 10 words that best describe their past emotions and then rate each emotion for how often it was experienced. This approach, the Emotion Recall Task (ERT), leverages on more effortful and accurate recall processes and avoids many concerns that surround an emotion checklist such as its breadth and specificity. Analysis showed that comparing with the PANAS, arguably the most commonly used emotional scale, the ERT performs at least equally well in predicting related constructs such as well-being, life satisfaction, and depression.

Chapter 3 examined the differences in performance between the norms-based ERT (ERT 1.0) and the self-rated ERT (ERT 2.0). By asking participants to rate the pleasantness and the arousal of their personal emotions—as opposed to relying on secondary valence and arousal norms—the ERT 2.0 offers a rapid approach to emotional measurement. We compared the ability of the ERT 2.0 to predict emotions on a set of additional scales: Diener, ONS, and BDI. Results showed that the ERT 2.0 better predicts these scales than the norms-based version on all constructs for both negative and positive emotions. PANAS, which we also compared, only outperforms the ERT 2.0 for negative emotions.

The results are promising because the ERT 2.0 allows individuals to produce words of any kind (even non-emotional words or words that are not in existing valence norms) and thus makes ERT 2.0 more specific to participants recalled emotions without the need to filter them through secondary norms. Secondly, the results are promising for cross-cultural comparisons and developmental studies because a well-developed knowledge of English is not necessary to understand what emotional words mean. Thirdly, the self-rated version is easy to score even without access to a computer and therefore may be valuable in situations where quick results are needed outside the laboratory.

The failure to predict negative emotions suggests a potentially distinct difference between recalled and recognized emotions. Why do people fail to recall negative emotions that are predictive of anxiety, even though they can easily recognize having had these emotions? This is of course an open question, but it interesting to speculate that it may be due to adaptive recall processes, that help individuals to show more positive emotion despite stressful conditions. In other words, it is potentially a form of adaptive self-deception (Trivers, 2000).

Chapter 4 aimed to examine the test–retest reliability of the ERT. Results revealed that the ERT is a reliable instrument for deriving recalled emotion with Spearman rank correlations on par with other comparable scales (e.g., PANAS, SWLS, ONS, and BDI).

Chapter 5 used an interview technique to gain additional insight into the emotional search process. When mentioning their recently experienced emotions, people try to retrieve information based on the most attentive situation they lately felt. It may be related to something in the near future or something in the past. Once they found an event, people selected one or more associates of that event

that evoked their emotions based its level of importance. This appears to follow past work suggesting that emotions are associated with events. It is also consistent with past work on memory search as a foraging process, but future work is needed here to understand the complete nature of this process. The emotional literacy data from Chapter 7 should be particularly useful for that investigation.

Chapter 6 investigated the applicability of the ERT 2.0 to an Indonesian sample and compared the emotional expression of an Indonesian sample with an American MTurk sample. Since the ERT 2.0 allows participants in any language to produce whatever words they like and is therefore content neutral with respect to emotional expectations, the ERT is a valuable instrument for studying emotional states across cultures. Our results suggest that American participants tend to produce more highly positive valence words, which was not expected based on inferences from Minangkabau culture (Kato, 1982). This is a topic that would be ideal for future study in Indonesia, as it could be examined by looking at age-related differences in emotional expression in Indonesian culture and also examining between ethnic groups with more or less exposure to globalization. Additional research was suggested to look at the ERT scores across a broader sample of countries. Finally, this chapter highlight the preliminary observation that negative affect in Indonesia may be related to past-oriented reflections on emotional state, that may be related to economic outcomes.

Chapter 7 provides additional exploration on how the mental lexicon and peak and end experiences influence recalled emotions. To complement one's recall on recently experienced emotions, participants were asked to recall as many emotion words as they could (not necessarily from their experience). This procedure is adapted from the concept of Verbal Fluency Task and involved a free recall method like the ERT, which we called *the emotional literacy task*. Results shows the ERT was not constrained by the number of emotions that one produced in the emotional literacy task. However, peak and end emotions both reflected a positive relationship with the ERT, explaining approximately half of its variance. This suggests that even our recent emotional experiences are governed by peak and end effects, implying they are also subject to duration neglect.

In sum, this work explored the relationship between recalled and recognized emotions and further explored what influences the emotions that we recall. Together, these results suggest a strong future for the ERT.

.

References

Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole.

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American journal of psychology*, 389-410.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological review, 79*(2), 97.

Anwar, K. (1980). *Indonesian: The development and use of a national language*. Gadjah Mada University Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation, 2*, 89-195.

Badan Pengembangan dan Pembinaan Bahasa, (2018, October). Kamus Besar Bahasa Indonesia (KBBI). Retrieved from https://kbbi.kemdikbud.go.id/

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological review, 77*(3), 215.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & emotion, 12*(4), 579-599.

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of personality and social psychology, 87*(2), 266.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & emotion, 15*(6), 713-724.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in cognitive sciences*, *11*(8), 327-332.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490-498.

Bock, M., & Klinger, E. (1986). Interaction of emotion and cognition in word recall. *Psychological Research*, *48*(2), 99-106.

Bower, G. H. (1981). Mood and memory. *American psychologist*, *36*(2), 129.

Bradburn, N. M. (1969). The structure of psychological well-being.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (2000). Emotion and motivation. *Handbook of psychophysiology*, *2*, 602-642.

Briggs, C. L. (1986). Learning how to ask: A sociolinguistic appraisal of the role of the interview in social science research(Vol. 1). Cambridge University Press.

Broadie, S., & Rowe, C. (2002). Aristotle: Nicomachean Ethics: Translation, Introduction, Commentary. Oxford: Oxford University Press.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*(1), 73-99.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: what is it, and how is it measured?. *Physiotherapy*, *86*(2), 94-99.

Carvalho, H. W. d., Andreoli, S. B., Lara, D. R., Patrick, C. J., Quintana, M. I., Bressan, R. A., . . . Jorge, M. R. (2013). Structural validity and reliability of the Positive and Negative Affect Schedule (PANAS): evidence from a large Brazilian community sample. *Revista Brasileira de Psiquiatria, 35*(2), 169-172.

Cella, D. F., Jacobsen, P. B., Orav, E. J., Holland, J. C., Silberfarb, P. M., & Rafla, S. (1987). A brief POMS measure of distress for cancer patients. *Journal of chronic diseases, 40*(10), 939-942.

Claeys, W., De Boeck, P., Van Den Bosch, W., Biesmans, R., & Böhrer, A. (1985). A comparison of one free-format and two fixed-format self-report personality assessment methods. *Journal of Personality and Social Psychology*, *49*(4), 1028.

Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology*, *63*(3), 352.

Clore, G. L., & Tamir, M. (2002). Affect as embodied information. *Psychological Inquiry, 13*(1), 37-45.

Clore, G. L., Gasper, K., Garvin, E., & Forgas, J. P. (2001). Affect as information. *Handbook of affect and social cognition*, 121-144.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. *Handbook of social cognition*, *1*, 323-417.

Crocker, P. R. (1997). A confirmatory factor analysis of the Positive Affect Negative Affect Schedule (PANAS) with a youth sport sample. *Journal of Sport & Exercise Psychology*.

Crouch, M., & McKenzie, H. (2006). The logic of small samples in interview-based qualitative research. *Social science information*, *45*(4), 483-499.

Darwin C. (1872). The Expression of the Emotions in Man and Animals. London: Murray.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71-75.

Diener, E., Napa-Scollon, C. K., Oishi, S., Dzokoto, V., & Suh, E. M. (2000). Positivity and the construction of life satisfaction judgments: Global happiness is not the sum of its parts. *Journal of happiness studies, 1*(2), 159-176.

Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D. W., & Oishi, S. (2009). New measures of well-being. *Assessing well-being*, 247-266.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143-156.

Ebesutani, C., Okamura, K., Higa-McMillan, C., & Chorpita, B. F. (2011). A psychometric analysis of the Positive and Negative Affect Schedule for Children–Parent Version in a school sample. *Psychological assessment, 23*(2), 406.

Edgeworth, F. Y., 1881. *Mathematical Psychics*. London: C. K. Paul.

Ehrhardt, J. J., Saris, W. E., & Veenhoven, R. (2000). Stability of life-satisfaction over time. *Journal of Happiness Studies*, *1*(2), 177-205.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and emotion*, *10*(2), 159-168.

Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and emotion*, *12*(3), 303-308.

Estes, W., & DaPolito, F. (1967). Independent variation of information storage and retrieval processes in paired-associate learning. *Journal of experimental psychology, 75*(1), 18-26.

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology, 69*(1), 153.

Fujita, F., & Diener, E. (2005). Life satisfaction set point: stability and change. *Journal of personality and social psychology*, *88*(1), 158.

Grove, J. R., & Prapavessis, H. (1992). Preliminary evidence for the reliability and validity of an abbreviated Profile of Mood States. *International Journal of Sport Psychology*.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods*, *18*(1), 59-82.

Guillet, R., & Arndt, J. (2009). Taboo words: The effect of emotion on memory for peripheral information. *Memory & Cognition*, *37*(6), 866-879.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255-282.

Hansen, C. H., & Shantz, C. A. (1995). Emotion-specific priming: Congruence effects on affect and recognition across negative emotions. *Personality and Social Psychology Bulletin*, *21*(6), 548-557.

Hebb, D. O. (1949). The organization of behavior. A neuropsychological theory.

Heider, K. G. (2006). Landscapes of emotion: Mapping three cultures of emotion in Indonesia. Cambridge University Press.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American sociological review*, 93-101.

Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. *Sociological methodology*, *2*, 104-129.

Helliwell, J., Layard, R., & Sachs, J. (2018). World Happiness Report 2018, New York: Sustainable Development Solutions Network.

Heppell, D. J. (2004). Penyebab dan Akibat Perubahan Kebudayaan Jawa di Yogyakarta. *Universitas Muhammdiyah Malang*.

Hills, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive science*, *30*(1), 3-41.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological review*, *119*(2), 431.

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science*, *7*(3), 513-534.

Hirst, W., Phelps, E. A., Meksin, R., Vaidya, C. J., Johnson, M. K., Mitchell, K. J., ... & Mather, M. (2015). A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events. *Journal of Experimental Psychology: General*, *144*(3), 604.

Hollingworth, H. L. (1913). Characteristic differences between recall and recognition. *The American Journal of Psychology, 24*(4), 532-544.

Huebner, E. S., & Dew, T. (1995). Preliminary validation of the positive and negative affect schedule with adolescents. *Journal of Psychoeducational Assessment, 13*(3), 286-293.

Kahneman, D. (2011). *Thinking, Fast and Slow. Book.* New York: Farrar, Straus and Giroux.

Kato, T. (1982). Matriliny and migration: Evolving Minangkabau traditions in Indonesia.

Kensinger, E. A., & Corkin, S. (2003a). Effect of negative emotional content on working memory and long-term memory. *Emotion*, *3*(4), 378.

Kensinger, E. A., & Corkin, S. (2003b). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words?. *Memory & cognition*, *31*(8), 1169-1180.

Killgore, W. D. S. (2000). Evidence for a third factor on the Positive and Negative Affect Schedule in a college student sample. *Perceptual and motor skills, 90*(1), 147-152.

Kintsch, W. (1970). Learning, memory, and conceptual processes. New York: Wiley.

Koriat, A. (2000). Control processes in remembering. *The Oxford handbook of memory*, 333-346.

Kornhuber, H. H. (1988). The human brain: from dream and cognition to fantasy, will, conscience and freedom. *Information processing by the brain. Toronto, Bern, Stuttgart: Hans Huber Publ*, 241-258.

Kuppens, P., Realo, A., & Diener, E. (2008). The role of positive and negative emotions in life satisfaction judgment across nations. *Journal of personality and social psychology*, *95*(1), 66.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 39-58.

Laurent, J., Catanzaro, S. J., Joiner Jr, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., . . . Gathright, T. (1999). A measure of positive and negative affect for children: scale development and preliminary validation. *Psychological assessment, 11*(3), 326.

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American psychologist*, *37*(9), 1019.

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of personality and social psychology*, *62*(6), 972.

Levy, R. I. (1984). The emotions in comparative perspective. *Approaches to emotion*, 397-412.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological Assessment* (4th ed.). New York: Oxford University Press.

Lim, N. (2016). Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integrative medicine research*, *5*(2), 105-109.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy*, *33*(3), 335-343.

Lucas, R. E., Diener, E., & Larsen, R. J. (2003). Measuring positive emotions. In S. J. Lopez & C. R. Snyder (Eds.),*Positive psychological assessment: A handbook of models and measures* (pp. 201–218). Washington, DC: American Psychological Association.

Lutz, C. (1988). Ethnographic perspectives on the emotion lexicon. In *Cognitive perspectives on emotion and motivation*(pp. 399-419). Springer, Dordrecht.

Lutz, C., & White, G. M. (1986). The anthropology of emotions. *Annual review of anthropology*, *15*(1), 405-436.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. Memory & Cognition, 32, 474-488.

Markowitsch, H. J., Kessler, J., Russ, M. O., Frölich, L., Schneider, B., & Maurer, K. (1999). Mnestic block syndrome. *Cortex*, *35*(2), 219-230.

McDowell, I. A. N., & Praught, E. D. (1982). On the measurement of happiness: an examination of the Bradburn Scale in the Canada Health Survey. *American journal of epidemiology*, *116*(6), 949-958.

McNair, P., Lorr, M., & Droppleman, L. (1981). POMS manual Educational and Industrial Testing Service. *San Diego*.

Merz, E. L., Malcarne, V. L., Roesch, S. C., Ko, C. M., Emerson, M., Roma, V. G., & Sadler, G. R. (2013). Psychometric properties of Positive and Negative Affect Schedule (PANAS) original and short forms in an African American community sample. *Journal of affective disorders, 151*(3), 942-949.

Mihić, L., Novović, Z., Čolović, P., & Smederevac, S. (2014). Serbian adaptation of the Positive and Negative Affect Schedule (PANAS): Its facets and second-order structure. *Psihologija, 47*(4), 393-414.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mueller, K. D., Koscik, R. L., LaRue, A., Clark, L. R., Hermann, B., Johnson, S. C., & Sager, M. A. (2015). Verbal fluency and early memory decline: results from the Wisconsin Registry for Alzheimer's prevention. *Archives of Clinical Neuropsychology*, *30*(5), 448-457.

Murdock, B. B. (1963). Short-term memory and paired-associate learning. *Journal of Memory and Language*, *2*(4), 320.

Noguchi, T., Stewart, N., Olivola, C. Y., Moat, H. S., & Preis, T. (2014). Characterizing the time-perspective of nations with search engine query data. *PLoS One*, *9*(4), e95209.

Oishi, S., Schimmack, U., & Colcombe, S. J. (2003). The contextual and systematic nature of life satisfaction judgments. *Journal of Experimental Social Psychology*, *39*(3), 232-247.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological bulletin*, *49*(3), 197.

Ostir, G. V., Smith, P. M., Smith, D., & Ottenbacher, K. J. (2005). Reliability of the Positive and Negative Affect Schedule (PANAS) in medical rehabilitation. *Clinical Rehabilitation, 19*(7), 767-769.

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010).

Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, *38*(4), 725-743.

Pavot, W., & Diener, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social Indicators Research*, *28*(1), 1-20.

Pavot, W., & Diener, E. (2009). Review of the satisfaction with life scale. In *Assessing well-being* (pp. 101-117). Springer, Dordrecht.

Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *Journal of marketing research*, 6-17.

Postman, L. (1963). *One-trial learning.* Paper presented at the Conference on Verbal Learning and Verbal Behavior, 2nd, Jun, 1961, Ardsley-on-Hudson, NY, US.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological review*, *88*(2), 93.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, *104*(1-2), 187-194.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin, 128*(6), 934.

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition & Emotion*, *5*(3), 161-200.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review, 110*(1), 145.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology, 76*(5), 805.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, *57*(6), 1069.

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of personality and social psychology*, *69*(4), 719.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*(3), 185-211.

Sapolsky, R. M. (1996). Why stress is bad for your brain. *Science*, *273*(5276), 749-750.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social science information, 44*(4), 695-729.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, *66*(2), 310-328.

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of personality and social psychology*, *89*(3), 395.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, *9*(5), 340-346.

Scollon, C. N., Diener, E., Oishi, S., & Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of cross-cultural psychology*, *35*(3), 304-326.

Shacham, S. (1983). A shortened version of the Profile of Mood States. *Journal of personality assessment, 47*(3), 305-306.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in psychology*, *5*, 772.

Shaver, P. R., Murdaya, U., & Fraley, R. C. (2001). Structure of the Indonesian emotion lexicon. *Asian journal of social psychology*, *4*(3), 201-224.

Sneddon, J. N. (2003). The Indonesian language: Its history and role in modern society. Sydney: UNSW Press.

Sowden, P. T., & Schyns, P. G. (2006) Channel surfing in the visual brain. Trends Cogn. Sci. 10, 538–545

Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., & Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *Journal of counseling psychology*, *49*(3), 381.

Stolwyk, R., Bannirchelvam, B., Kraan, C., & Simpson, K. (2015). The cognitive abilities associated with verbal fluency task performance differ across fluency variants and age groups in healthy young and old adults. *Journal of clinical and experimental neuropsychology*, *37*(1), 70-83.

Suraya. 2003. Peranan Komunikasi dalam Penyatuan Budaya. Jurnal Universitas Paramadina. Vol. 3, No.1, hal. 124 - 135.

Tabor, D., & Stockley, L. (2018, May 17). Personal well-being in the UK: January to December 2017. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/measuringnationalwellbeing/januarytodecember2017

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological science*, *14*(5), 455-461.

Terry, P. C., Lane, A. M., Lane, H. J., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. *Journal of sports sciences, 17*(11), 861-872.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of personality and social psychology*, *59*(2), 291.

Thompson, E. R. (2007). "Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS)". *Journal of Cross-Cultural Psychology*. 38 (2): 227–242. doi:10.1177/0022022106297301. ISSN 0022-0221.

Tomkins, S. S. (1984). Affect theory. *Approaches to emotion*, *163*(163–195).

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, *907*(1), 114-131.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *neuropsychology*, *11*(1), 138.

Tugade, M. M., Fredrickson, B. L., & Feldman Barrett, L. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of personality, 72*(6), 1161-1190.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*(2), 175-184.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 381-391.

Van Rensbergen, B., Kuppens, P., Storms, G., & De Deyne, S. (2015). Computationally coding responses of a free-format self-description personality test using word association data. Manuscript submitted for publication.

Van Rensbergen, B., Storms, G., & De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic bulletin & review*, *22*(6), 1717-1724.

Verduyn, P., & Lavrijsen, S. (2015). Which emotions last longest and why: The role of event importance and rumination. *Motivation and Emotion*, *39*(1), 119-127.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, *5*(4), 672-688.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, *45*(4), 1191-1207.

Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 687-704.

Watson, D. (2000). *Mood and temperament*: Guilford Press.

Watson, D., & Clark, L. A. (1994). Manual for the positive and negative affect schedule (expanded form). *Unpublished manuscript, University of Iowa, Iowa City*.

Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063.

Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: psychometric patterns and behavioral and physiological responses to stress. *Journal of abnormal psychology*, *88*(4), 369.

Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very long-term memory. *Cognitive science, 5*(2), 87-119.

Wilson, K., Gullone, E., & Moss, S. (1998). The youth version of the Positive and Negative Affect Schedule: A psychometric validation. *Behaviour Change, 15*(03), 187-193.

Winskel, H., & Widjaja, V. (2007). Phonological awareness, letter knowledge, and literacy development in Indonesian beginner readers and spellers. *Applied Psycholinguistics*, *28*(1), 23-45.

Wong, C. S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The leadership quarterly*, *13*(3), 243-274.

Wundt, W. (1905a). Fundamentals of physiological psychology. *Leipzig: Engelmann*.

Wundt, W. M. (1905b). Völkerpsychologie: eine untersuchung der entwicklungsgesetze von sprache, mythus und sitte (*Peoples psychology: an investigation of the laws of evolution of language, myth and manners*) (Vol. 2): W. Engelmann.

Wyer, R. S., Clore, G. L., & Isbell, L. M. (1999). Affect and information processing. *Advances in experimental social psychology, 31*, 1-77.

Wyrick, R. A., & Wyrick, L. C. (1977). Time experience during depression. *Archives of General Psychiatry*, *34*(12), 1441-1443.

Wysokiński, A., Zboralski, K., Orzechowska, A., Gałecki, P., Florkowski, A., & Talarowska, M. (2010). Normalization of the Verbal Fluency Test on the basis of results for healthy subjects, patients with schizophrenia, patients with organic lesions of the chronic nervous system and patients with type 1 and 2 diabetes. *Archives of medical science: AMS*, *6*(3), 438.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of applied psychology*, *71*(2), 291.

Zemla, J. C., & Austerweil, J. L. (2018). Estimating semantic networks of groups and individuals from fluency data. *Computational Brain & Behavior*, 1-23.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, *43*(1), 111.

Appendix A

THE EMOTIONAL RECALL TASK - English Version

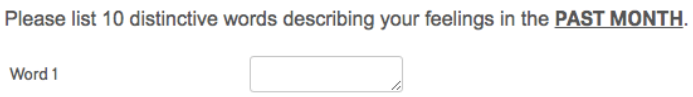
Section 1: Consent
INFORMATION SHEET
DEPARTMENT OF PSYCHOLOGY - UNIVERSITY OF WARWICK

Dear Participant,
You are invited to a research study assessing emotions and life quality.

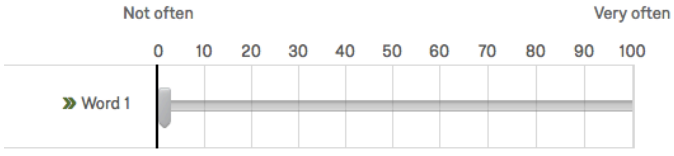
What will happen
You will be asked to fill in a set of questionnaires, and you will also be required to provide general demographic information. This study takes approximately 30 minutes to complete. No deception is used in this survey, and there are no right or wrong answers.

Participants' rights
Your participation in this study is completely voluntary. You are free to terminate your participation at any time by closing the web tab. We take participants' privacy very seriously. Specifically, no personal identification questions will be collected and your answers will be completely anonymous. If you have any questions with this information sheet, you should clarify with the researcher before you start. I can be reached at Masitah.-@warwick.ac.uk.

Section 2: ERT Emotions
"Please list 10 distinctive words describing your emotions in the PAST MONTH"



Section 3: ERT Amounts
In this page, we will show you all the words you just used to describe your feelings. Please use the slider bar to tell us how often you have experienced each of these emotions in the PAST MONTH.



Appendix B

THE EMOTIONAL RECALL TASK – ENGLISH NORMS Chapter 3

Word	n	V.Mean	V. SD	V.Norm	A.Mean	A. SD	A.Norm
accomplished	3	9.67	0.58	0.66	5.33	1.53	1.73
active	5	7.80	1.30	1.14	4.60	1.14	1.00
afraid	7	5.29	3.68	2.73	7.29	1.50	1.11
aggressive	2	6.50	2.12	2.94	9.00	0.00	0.00
alert	6	7.17	2.32	1.85	6.17	2.23	1.78
alive	2	7.00	1.41	1.96	7.00	2.83	3.92
ambitious	4	7.25	3.10	3.03	7.75	1.71	1.67
amused	7	8.71	1.38	1.02	6.71	1.11	0.82
angry	21	4.38	2.96	1.26	6.52	2.09	0.89
annoyed	6	3.00	2.28	1.82	6.67	2.07	1.65
apathetic	1	8.00	0.00	0.00	5.00	0.00	0.00
appreciated	1	10.00	0.00	0.00	2.00	0.00	0.00
assertive	2	7.50	0.71	0.98	7.00	1.41	1.96
attractive	3	8.00	2.00	2.26	6.33	3.79	4.28
happy	104	8.30	1.91	0.37	6.15	2.55	0.49
humorous	1	10.00	0.00	0.00	10.00	0.00	0.00
hungover	1	4.00	0.00	0.00	1.00	0.00	0.00
hungry	9	4.33	2.87	1.88	6.11	2.26	1.48
hurt	2	3.50	3.54	4.90	8.00	0.00	0.00
ill	1	1.00	0.00	0.00	5.00	0.00	0.00
lazy	5	5.20	2.39	2.09	1.60	0.55	0.48
lifeless	1	1.00	0.00	0.00	5.00	0.00	0.00
lonely	10	2.50	2.27	1.41	3.90	2.08	1.29
longing	1	1.00	0.00	0.00	3.00	0.00	0.00
lost	1	2.00	0.00	0.00	3.00	0.00	0.00
love	10	6.60	1.71	1.06	5.80	2.44	1.51
loyal	1	9.00	0.00	0.00	7.00	0.00	0.00
lucky	3	8.33	1.15	1.31	6.67	2.52	2.85
mad	2	7.00	0.00	0.00	5.00	1.41	1.96
overwhelmed	4	3.00	2.71	2.65	8.25	1.71	1.67
pain	4	3.00	4.00	3.92	7.75	2.63	2.58
painful	1	1.00	0.00	0.00	6.00	0.00	0.00
panic	1	7.00	0.00	0.00	7.00	0.00	0.00
passionate	2	10.00	0.00	0.00	10.00	0.00	0.00
romantic	2	9.00	1.41	1.96	4.50	3.54	4.90
rushed	4	2.25	1.50	1.47	9.00	1.41	1.39
sad	52	4.50	3.15	0.86	5.58	2.55	0.69
sadness	6	6.17	2.99	2.40	6.67	2.88	2.30
safe	2	6.00	2.83	3.92	4.00	4.24	5.88
satisfied	12	8.75	1.22	0.69	4.50	3.15	1.78
scared	3	3.33	4.04	4.57	7.33	1.15	1.31
secure	1	9.00	0.00	0.00	1.00	0.00	0.00
selfish	1	5.00	0.00	0.00	5.00	0.00	0.00
sleepy	5	3.20	1.92	1.69	3.40	2.51	2.20
sluggish	1	1.00	0.00	0.00	1.00	0.00	0.00
straightforward	1	8.00	0.00	0.00	7.00	0.00	0.00
useless	1	1.00	0.00	0.00	3.00	0.00	0.00
worried	20	1.85	1.39	0.61	6.75	2.12	0.93

Appendix C

THE EMOTIONAL RECALL TASK – INTERVIEW MANUAL

Section 1: Rapport and Consent
"You are invited to a research study on emotional literacy. We especially interested in the way people express their recent experienced emotions.
I am Masitah, from Department of Psychology at the University of Warwick. This project is supervised by Professor Thomas Hills, and *it has been approved by the Psychology Research Ethics Committee*.

You will be asked to fill in *paper-task and interview-task*. This study takes approximately *20 minutes to complete*. No deception is used in this survey, and there are no right or wrong answers.

Your participation in this study is completely voluntary. You are free to terminate your participation at any time. We take participants' privacy very seriously. Specifically, no personal identification questions will be collected and your answers will be completely anonymous. If you have any questions with this information sheet, you should clarify with the researcher before you start. "

Section 2: Paper Tasks
Participant was asked to write the 10 emotions during the past month on the paper and rate each of the emotions on a scale from 0 to 100.

Section 3: Interview Tasks
Part 1: Interviewer probing each of the emotions. "Before you said this (particular) word, can you explain what is come to your mind?"

Part 2: "Is it really come to your mind before you said this (particular) word, or you just thinking about it now because I am asking?"

Part 3: Are you consciously get this for **the past month** (time frame)?

Part 4: Participant asked to rate the emotion again**. "**You rated each of those emotions. If I ask you to rate it again on a scale from 0 to 100, how frequent do you have experienced each of those emotions in the last past month?"

Section 4: Debriefing
"This study aims to understand the relationship between people's subjective wellbeing by looking at how they recall emotions.
Thank you for participating."

Appendix D

THE EMOTIONAL RECALL TASK - Indonesian Version

Section 1: Consent

Bapak/Ibu yang saya hormati,

Saya Annasya Masitah, peneliti dari Departement Psikologi University of Warwick - Inggris. Saya sedang mengerjakan Cultural Project terkait metode pengukuran emosi, dan tertarik mempelajari lebih dalam mengenai cara seseorang mengekspresikan apa yang sedang dirasakan. Cultural Project ini dilakukan di beberapa negara, termasuk Indonesia. Saat ini Anda secara khusus mendapat kesempatan untuk berpartisipasi.

Kuesioner pada *Cultural Project* ini terdiri dari 5 (lima) bagian dan Anda diminta untuk memilih atau memberikan jawaban singkat untuk setiap pertanyaan mengenai **apa yang Anda rasakan dalam jangka waktu satu bulan terakhir.** Kuesioner ini tidak menilai jawaban benar atau salah, dan tidak menggunakan manipulasi apapun. Partisipan umumnya menyelesaikan kuesioner dalam waktu 15-20 menit.

Jika ada yang ingin ditanyakan terkait kuesioner ini, saya dapat dihubungi melalui email: Masitah.-@warwick.ac.uk

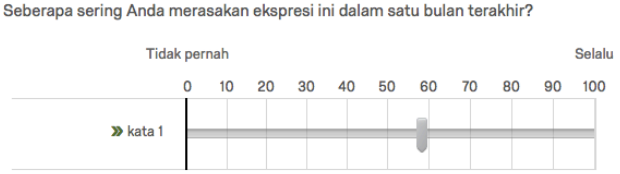
Section 2: ERT Emotions

"Silahkan menulis 10 kata berbeda yang mengekspresikan **apa yang Anda rasakan dalam satu bulan terakhir**."

Section 3: ERT Amounts

Bagian berikut berisi ekspresi perasaan yang Anda isikan pada bagian awal.

Anda diharapkan menggunakan skala 0-100 untuk menunjukkan frekuensi yang Anda rasakan terhadap masing-masing ekspresi tersebut dalam satu bulan terakhir.



Appendix E

THE EMOTIONAL RECALL TASK – INDONESIAN NORMS Chapter 6

Indonesian Word	English Word	n	V.Mean	V. SD	V.Norm	A.Mean	A. SD	A.Norm
abai	neglected	3.00	2.67	2.08	2.36	4.67	2.89	3.27
abstrak	abstract	1.00	5.00	0.00	0.00	2.00	0.00	0.00
acuh	indifferent	7.00	4.14	2.79	2.07	5.71	3.35	2.48
adaptasi	adaptation	1.00	5.00	0.00	0.00	3.00	0.00	0.00
adem	cool	2.00	8.00	2.83	3.92	6.50	4.95	6.86
adik	younger brother / sister	1.00	10.00	0.00	0.00	5.00	0.00	0.00
agamis	concerning belief in divinity	1.00	9.00	0.00	0.00	9.00	0.00	0.00
air mata	tears	3.00	3.67	3.79	4.28	7.00	2.00	2.26
akhirat	afterlife	1.00	9.00	0.00	0.00	7.00	0.00	0.00
alhamdulillah	grateful in Islam	5.00	6.40	3.51	3.07	8.60	2.19	1.92
Allah	Allah	6.00	8.83	1.83	1.47	7.83	2.64	2.11
asa	hope -like expression	2.00	4.00	2.83	3.92	8.50	2.12	2.94
asyik	preoccupied	3.00	6.00	4.58	5.19	6.67	3.51	3.97
baper	hard feeling	6.00	4.83	2.56	2.05	6.00	3.85	3.08
begadang	stay up late	2.00	2.50	2.12	2.94	4.50	0.71	0.98
bengong	to gaze emptily in silence as if you've lost your mind	1.00	1.00	0.00	0.00	7.00	0.00	0.00
berada	well-off	2.00	7.00	4.24	5.88	6.00	0.00	0.00
berantakan	messy	3.00	1.00	0.00	0.00	6.33	3.21	3.64
berapa	how many	1.00	6.00	0.00	0.00	7.00	0.00	0.00
berbunga	flowery	1.00	7.00	0.00	0.00	7.00	0.00	0.00
berbunga-bunga	blossomed	2.00	8.00	1.41	1.96	4.00	4.24	5.88
berdebar	pounding	3.00	4.33	3.06	3.46	5.33	2.89	3.27
gaji buta	Getting paid but doing nothing	1.00	5.00	0.00	0.00	4.00	0.00	0.00
gemas	a mix feelings of intense love and hate	2.00	6.50	0.71	0.98	4.00	4.24	5.88
gereget	being angry/annoyed in a playful way	1.00	10.00	0.00	0.00	9.00	0.00	0.00
geregetan	a feeling that something lacks a certain unnameable thing	3.00	3.67	4.62	5.23	5.67	0.58	0.65
gotong-royong	mutual cooperation	1.00	8.00	0.00	0.00	2.00	0.00	0.00
haru	Indonesian mix feeling expression	23.00	6.26	2.26	0.92	5.30	2.36	0.97
heboh	hysterical	2.00	4.50	2.12	2.94	5.00	0.00	0.00
itikaf	stayed up all night for worshipping	1.00	10.00	0.00	0.00	8.00	0.00	0.00
jantungan	suffer a heart attack	1.00	5.00	0.00	0.00	6.00	0.00	0.00
jatuh	falling down	5.00	5.80	3.27	2.87	6.40	2.88	2.53
jatuh cinta	in love	5.00	5.00	3.16	2.77	4.20	3.35	2.93
jenuh	saturated feeling	53.00	4.77	3.35	0.90	5.94	2.74	0.74
kepikiran	keep thinking about it	2.00	1.00	0.00	0.00	5.50	0.71	0.98
keponakan	nephew/niece	1.00	10.00	0.00	0.00	10.00	0.00	0.00
khilaf	make a mistake	1.00	3.00	0.00	0.00	4.00	0.00	0.00
malah	on the contrary	1.00	3.00	0.00	0.00	6.00	0.00	0.00
mandiri	be autonomous	2.00	7.00	1.41	1.96	6.50	2.12	2.94
mati rasa	numbness	1.00	9.00	0.00	0.00	9.00	0.00	0.00
melupakan	take one's mind off	2.00	5.50	6.36	8.82	8.00	0.00	0.00
membisu	tight-lipped	1.00	1.00	0.00	0.00	2.00	0.00	0.00
memudahkan	make it easy	1.00	1.00	0.00	0.00	10.00	0.00	0.00
menderita	suffer from	3.00	2.67	2.89	3.27	7.00	0.00	0.00
mendongkol	crank up	1.00	2.00	0.00	0.00	4.00	0.00	0.00
menduga	guessed	1.00	2.00	0.00	0.00	1.00	0.00	0.00
mendukung	support	1.00	8.00	0.00	0.00	7.00	0.00	0.00
menegangkan	tense	3.00	5.33	4.04	4.57	6.00	1.73	1.96
menempuh	go through	1.00	1.00	0.00	0.00	5.00	0.00	0.00
menenangkan	calm down	1.00	10.00	0.00	0.00	1.00	0.00	0.00
mengasihi	be attached	1.00	7.00	0.00	0.00	7.00	0.00	0.00
menyesal	feel sorry for	34.00	3.71	3.12	1.05	6.21	2.24	0.75
miris	feeling of mercy toward another	1.00	2.00	0.00	0.00	6.00	0.00	0.00
pada	in the direction of	1.00	4.00	0.00	0.00	3.00	0.00	0.00
pamrih	strings attached	1.00	10.00	0.00	0.00	9.00	0.00	0.00
pemurung	downhearted	1.00	2.00	0.00	0.00	3.00	0.00	0.00
plinplan	wishy-washy	1.00	1.00	0.00	0.00	3.00	0.00	0.00
plong	feel free from the burden of the mind	1.00	8.00	0.00	0.00	7.00	0.00	0.00
sholat	prayer in Islam	1.00	9.00	0.00	0.00	7.00	0.00	0.00
tangkis	resist punches or attacks with weapons and so on	1.00	10.00	0.00	0.00	8.00	0.00	0.00
tilawah	Recitationthe extracts or portions from the Qur'an	1.00	10.00	0.00	0.00	5.00	0.00	0.00
tulus	heart-whole	7.00	7.57	2.70	2.00	7.14	2.19	1.62

Appendix F

THE FRAGMENTARY

Below is the description of ERT related projects that are still ongoing or could not be completed before the due date for this thesis. Some faced obstacles in the middle of experiments, some did not provide enough evidence to be wrapped up as a complete study, and some are still being carried out.

A. Emotional search and network analysis

This project is the followed up from Chapter 7 on the emotional literacy data. This project aims to explore the possibility that the network structure of the emotional lexicon can predict emotional recall. The analyses conducted include distinctiveness of emotions (network degree—number of neighbours) and relative clustering of emotions (network clustering coefficient—how well neighbours are connected). Network features (degree and clustering coefficient) were obtained through the network analysis of an adjacency-based fluency network where the nodes were emotions and the edges were the number of bi-directional consecutive generation of emotions. For example, *happy* produced after *sad* would lead to a directed edge leading from *happy* to *sad* in the network. This project will examine correlation analysis between network properties of emotions in the emotional literacy task and their production in the ERT.

B. Cultural projects: India and Malaysia

This project was run as a component of Chapter 7 but eventually failed due to logistical challenges. Before scheduling the data collection in Indonesia, the data collection process started in India. The India project started earlier after an undergraduate student from the Indian Institute of Technology Kharagpur contacted me to offer to help running the data collection for an Indian

population, however, we only received responses from 15 participants even after prolonging the time period and promoting more cash reward.

For the Malaysia project, the data collection process went pretty well. I received good responses from the University of Malaya – Kuala Lumpur, the University Pendidikan Sultan Idris – Tanjong Malim, and the University Teknologi Mara – Selangor. We collected responses from 107 participants. Unfortunately, because the survey was carried out in English (before the completion of the paper and pencil version), the data failed to capture the authenticity and uniqueness of the Melayu culture.